EKONOMETRIA ECONOMETRICS 4(54) • 2016

ISSN 1507-3866
e-ISSN 2449-9994

Justyna Brzezińska
University of Economics in Katowice
e-mail: justyna.brzezinska@ue.katowice.pl

ANALYSIS OF LATENT CLASS MODELS IN ECONOMIC RESEARCH

ANALIZA MODELI ZMIENNYCH UKRYTYCH W BADANIACH EKONOMICZNYCH

DOI: 10.15611/ekt.2016.4.02
JEL Classification: C51, C52, C87

Summary: Latent variable models are used and applied in many areas of the social and behavioral sciences. The increasing availability of computer packages for fitting such models makes latent variable models popular, known and applied in many scientific areas. Latent variable models have a very wide range of applications, especially in the presence of repeated observations, longitudinal data, and multilevel data. The basic model postulates an underlying categorical latent variable; within any category of the latent variable the manifest or observed categorical variables are assumed independent of one another (the axiom of conditional independence). The observed relationships between the manifest variables are thus assumed to result from the underlying classification of the data produced by the categorical latent variable. In this paper we present the theoretical and methodological aspects of latent variable models, as well as their application in R software in the field of economic research.

Keywords: latent class models, latent variables, categorical data analysis, R software.

Streszczenie: Modele zmiennych ukrytych są z powodzeniem stosowane w wielu obszarach badawczych, szczególnie w naukach społecznych. Wzrost dostępności nowoczesnych programów komputerowych pozwalających na budowę oraz dopasowanie złożonych modeli statystycznych sprawił, że modele te stają się coraz bardziej popularne w wielu innych dyscyplinach naukowych. Modele oparte na zmiennych ukrytych mają szerokie zastosowanie, szczególnie w przypadku analizy danych panelowych czy też wielopoziomowych. Podstawowy model bazuje na założeniu, że zmienna ukryta ma charakter jakościowy, a dla każdej kategorii zmiennej ukrytej jakościowe zmienne obserwowalne są niezależne od siebie (aksjomat lokalnej niezależności). W niniejszym artykule przedstawione zostaną teoretyczne i metodologiczne aspekty modeli opartych na zmiennych ukrytych wraz z ich zastosowaniem w naukach ekonomicznych przy wykorzystaniu programu R.

Słowa kluczowe: modele klas ukrytych, zmienne ukryte, analiza danych jakościowych, program R.

1. Introduction

The main development of latent class analysis took place during the second half of the twentieth-century. The practical application of these models by researchers in various fields of inquiry became a realistic possibility only in the last quarter of the twentieth century. Although the problem of measuring the relationships (the non-independence) between two or more observed dichotomous or polytomous data arose and was considered in many fields of inquiry at various times throughout the nineteenth and twentieth century, we can expect that researchers in some of these fields of inquiry, and in other fields as well, will find that the introduction and application of latent class models can help them to gain further insight into the observed relations among these observed variables of interest.

Latent variable models have gradually become an integral part of mainstream statistics and are currently used in a multitude of applications in different subject areas. Examples include, to name but a few, longitudinal analysis [Verbeke, Molenberghs 2000], covariate measurement error [Carroll i in. 2006], multivariate survival [Hougaard 2000], market segmentation [Wedel, Kamakura 2000], psychometric measurement [McDonald 1999], meta-analysis [Sutton 2000], capture–recapture [Coull, Agresti 1999], discrete choice [Train 2003], biometrical genetics [Neale, Cardon 1992] and spatial statistics [Rue, Held 2005]. Thirty-five years ago the Danish statistician Andersen published an important survey of latent variable modeling in the Scandinavian Journal of Statistics [Andersen 1982], in a paper entitled “Latent Structure Analysis: A Survey”.

Reference should also be made, of course, to many other publications by others, in the 1970s and later, that also contributed to the further development of this subject. With respect to the estimation procedures for latent class models, the reader is referred to, for example, Haberman [1976], Dempster et al. [1977)], and Vermunt [1997; 1999]. In addition to the reference to Haberman’s work (on estimation procedures for latent class models) we also refer the interested reader to Haberman [1974; 1979] for related material. With respect to the various reviews of the latent class models literature, which covered work done in the late 1970s and later, we cite here, for example, Clogg [1981], Clogg and Sawyer [1981], and Andersen [1982].

The literature on latent variable models that appeared in the 1950s and the 1960s was primarily limited to the situation in which all of the observed (manifest) variables were dichotomous (not polytomous). Lazarsfeld first introduced the term latent structure models in 1950, and the various models that he considered as latent structure models (including the latent class model) were concerned mainly with the “dichotomous systems” of observed variables.

During the 1950s and 1960s there were essentially five different methods that were proposed for estimating the parameters in the latent class model:

(1) a method suggested by Green [1951] that resembled in some respects a traditional factor analysis;

(2) a method suggested by Gibson [1955] that was quite different from the method suggested by Green [1951] but that also resembled factor analysis in some respects;
(3) a method that was based on the calculation of the solution of certain determinantal equations that was suggested by the work of Lazarsfeld and Dudman [1951] and Koopmans [1951] and that was developed by Anderson [1954] and extended by Gibson [1955] and Madansky [1960];
(4) a scoring method described by McHugh [1956] for obtaining maximum-likelihood estimates of the parameters in the model;
(5) a partitioning method developed by Madansky [1959] that is based on an examination of each of the possible assignments of the observations in the cross-classification table to the different latent classes.

Some work in these areas has focused on their application in social sciences [Sobel 1994; Edwards and Bagozzi 2000; Hägglund 2001; and Chung, Anthony, Shen 2009; Schafer 2011].

Latent variable models are used and applied in many areas of the social and behavioural sciences. The increasing availability of computer packages for fitting such models makes latent variable models popular, known and applied in many scientific areas. Latent variable models have a very wide range of applications, especially in the presence of repeated observations, longitudinal data, and multilevel data. The basic model postulates an underlying categorical latent variable; within any category of the latent variable the manifest or observed categorical variables are assumed independent of one another (the axiom of conditional independence). The observed relationships between the manifest variables are thus assumed to result from the underlying classification of the data produced by the categorical latent variable.

Different types of the latent variable model can be grouped according to whether the manifest and latent variables are categorical or continuous. When both the latent and manifest variables are continuous we use factor analysis. When the latent variable is categorical and the manifest variable is continuous, we use latent profile analysis. When the latent variable is continuous, and the manifest variable is categorical we use item response theory models. When both the manifest and latent variables are categorical, we use latent class analysis.

In this paper we present the theoretical and methodological aspects of latent variable models, as well as their application in R software in the field of economic research.

2. Latent variable models

Latent variable models form the latent variables used in the structural model. When a latent variable model is analyzed without a structural model, it is called a confirmatory factor analysis (CFA). If there was not a hypothesized structure for the latent variable model, then it would be an exploratory factor analysis (EFA).

Latent Class Analysis (LCA) is a statistical method used in factor analysis, cluster analysis, and regression methods. In this method, constructs are identified and created from unobserved, or latent, subgroups, which are usually based on individual responses from multivariate categorical data. Latent class analysis, along with latent trait analysis (discussed later), have their roots in the work of the sociologist, Paul Lazarsfeld, in the 1960s. Under the general methods of latent structure analysis these techniques were intended as tools of sociological analysis. Although Lazarsfeld recognized certain affinities with factor analysis he emphasized the differences. Thus in the old approach these families of methods were regarded as quite distinct. Although statistical theory had made great strides since Spearman's time there was little input from statisticians until Leo Goodman began to develop efficient methods for handling the latent class model around 1970. Although latent variables are not observable, some of their effects on measurable (manifest) variables are observable, and hence subject to study. Indeed, one of the major achievements in the behavioral sciences has been the development, over several decades, of methods to assess and explain the structure in a set of correlated, observed variables, in terms of a small number of latent variables. Latent variables occur in many areas, for example in psychology, intelligence and verbal ability, in sociology, ambition and racial prejudice, and in economics, economic expectation. Clearly, direct measurement of a concept such as racial prejudice is not possible, however, one could, for example, observe whether a person approves or disapproves of a particular piece of government legislation, whether the numbers of people of a particular race among his/her friends and acquaintances, etc., and assume that these are, in some sense, indicators of a more fundamental variable, racial prejudice. In some cases the manifest variables will be discrete (nominal) variables, in others continuous (interval or ratio) variables, and, as we shall see later, such a classification may also often be usefully applied to the latent variables.

3. The general model

The aim of many of the techniques to be described in this text is a simplified description of the structure of the observations by means of what is usually referred to as a model. This can range from a fairly imprecise verbal description to a geometrical representation or mathematical equation. The latter is a precise description of what the investigator visualizes as occurring in the population of interest and may, in many cases, provide the basis for a formal significance test of the model. The purpose of building a model is to provide the simplest explanation of the phenomena under investigation that is consistent with the observations; of course, if a model is made complex enough it is bound to provide an adequate fit, but a complicated model may have less explanatory power than one which is simpler but more elegant. Also, the simpler the model, the easier the task of interpretation.

Now let us consider the latent class model in a situation in which variable *A* is an observed (or manifest) dichotomous or polytomous variable having *I* classes ($i = 1, 2, \dots, I$), variable *B* is an observed (or manifest) dichotomous or polytomous variable having *J* classes ($j = 1, 2, \dots, J$), and variable *X* is an unobserved (or latent) dichotomous or polytomous variable having *T* classes ($t = 1, 2, \dots, T$).

Let π_{ijt}^{ABX} denote the joint probability that an observation is in class *i* on variable *A*, in class *j* on variable *B*, and in class *t* on variable *X*; let π_{it}^{AX} denote the conditional probability that an observation is in class *i* on variable *A*, given that the observation is in class *t* on variable *X*; let π_{jt}^{BX} denote the conditional probability that an observation is in class *j* on variable *B*, given that the observation is in class *t* on variable *X*; and let π_{t}^{X} denote the probability that an observation is in class *t* on variable *X*. The latent class model in this situation can be expressed simply as follows:

$$\pi_{ijt}^{ABX} = \pi_{t}^{X} \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X}, \tag{1}$$

for $i = 1, 2, \dots, I$, $j = 1, 2, \dots, J$, $t = 1, 2, \dots, T$.

This model states that variables *A* and *B* are conditionally independent of each other, given the class level on variable *X*. That is,

$$\pi_{ijt}^{\bar{A}\bar{B}X} = \pi_{ijt}^{ABX} / \pi_{t}^{X} = \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X}, \tag{2}$$

where $\pi_{ijt}^{ABX} = \pi_{ijt}^{ABX} / \pi_{t}^{X}$ is the conditional probability that an observation is in class *i* on variable *A* and in class *j* on variable *B*, given that the observation is in class *t* on variable *X*.

This situation is described when there are only two observed (manifest) variables (say, *A* and *B*). This we do for expository purposes in order to consider this subject in its simplest context. However, it should be noted that some special problems arise when latent class models are considered in a situation in which there are only two observed variables that do not arise in a situation in which there are more than two observed variables. However, these problems need not deter us here. For illustrative purposes, next we shall consider some examples in which latent class models are applied in the analysis of cross-classified data in the case in which there are two observed variables and also in the case where there are more than two observed variables.

An important part of the model selection procedure in latent class analysis involves checking whether a model is in agreement with the data. The discrepancies between observed data and expectations under the model can be assessed using goodness-of-fit (GoF) statistics. In this paper we present and apply statistics for the assessment of global and local fit.

Global fit statistics aggregate the disagreement between the observed frequencies and the expected frequencies under the model. One of the most popular and known statistic for testing the goodness-of-fit is Pearson's chi-square χ^2 and likelihood ratio statistics G^2, or information criteria such as AIC and BIC. In this paper we will use information criteria for model selection.

4. Application in R

Latent class analysis is a technique for the analysis of clustering among observations in multi-way tables of qualitative or categorical variables. The central idea of this method is to fit a model in which any confounding between the manifest variables can be explained by a single unobserved latent categorical variable. `poLCA` package available in R uses the assumption of local independence to estimate a mixture model of latent multi-way tables, the number of which (`nclass`) is specified by the user. Estimated parameters include the class-conditional response probabilities for each manifest variable, the mixing proportions denoting population share of observations corresponding to each latent multi-way table, and coefficients on any class-predictor covariates, if specified in the model.

`polLCA` uses EM and Newton-Raphson algorithms to maximize the latent class model log-likelihood function. Depending on the starting parameters, this algorithm may only locate a local, rather than global, maximum. This becomes more and more of a problem as `nclass` increases. It is therefore highly advisable to run `poLCA` multiple times until one is relatively certain that one has located the global maximum log-likelihood. As long as `probs.start=NULL`, each function call will use different (random) initial starting parameters. Alternatively, setting `nrep` to a value greater than one enables the user to estimate the latent class model multiple times with a single call to `poLCA`, thus conducting the search for the global maximizer automatically.

In this paper we present the application of latent class analysis for the analysis of dichotomous ratings by seven pathologists of 118 slides for the presence or absence of carcinoma in the uterine cervix. The data is available in R as `carcinoma` in `poLCA` package [Agresti 2002]. In this dataset, the pathologists are labeled `A` through `G`. There are 20 different observed response patterns. The data consists of 118 observations on 7 variables representing pathologist ratings with 1 denoting "no" and 2 denoting "yes".

Firstly, we fit the model for 2 latent classes. Conditional item response (column) probabilities, by outcome variable for each class (row) are presented in Table 1.

The estimated class population shares are: 0.5012 and 0.4988. The maximum log-likelihood for the model with 2 latent classes is –317.2568. The information criteria are the following: AIC = 664.5137, BIC = 706.0739, and likelihood ratio $G^2 = 62.3654$, and the chi-square goodness of fit statistic $\chi^2 = 92.6481$.

Figure 1 presents the probability of latent class membership for the model with 2 latent variables.

In the next part of this paper we build a model with 3 latent classes. The conditional item response (column) probabilities, by outcome variable for each class (row) are presented in Table 2.

Table 1. Conditional item response (column) probabilities, by outcome variable for each class (row) for a model for 2 latent classes

Variable	Pr(1)		Pr(2)	
	Class 1	Class 2	Class 1	Class 2
A	0.0000	0.8835	1.0000	0.1165
B	0.0169	0.6456	0.9831	0.3544
C	0.2391	1.0000	0.7609	0.0000
D	0.4589	1.0000	0.5411	0.0000
E	0.0214	0.7771	0.9786	0.2229
F	0.5773	1.0000	0.4227	0.0000
G	0.0000	0.8835	1.0000	0.1165

Source: own calculations in R based on Agresti [2002].

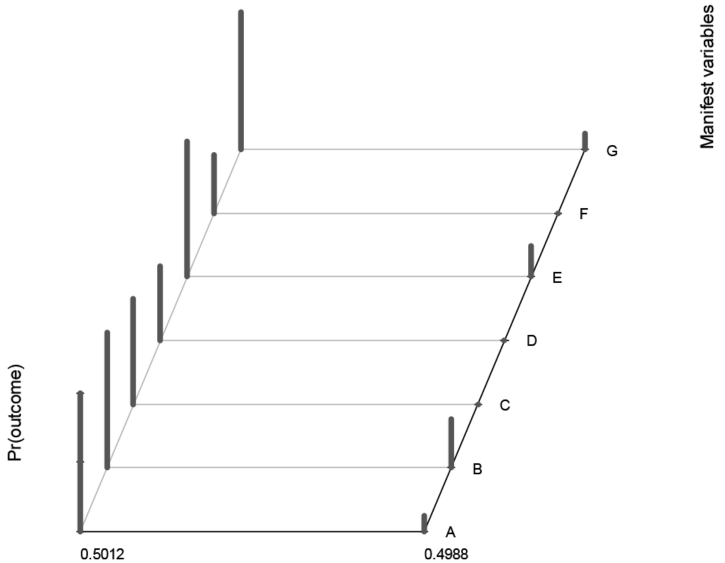

Fig. 1. Probability of latent class membership for model for 2 latent variables

Source: own calculations in R.

The estimated class population shares are: 0.3736, 0.1817, and 0.4447. The maximum log-likelihood for the model with 3 latent classes is –293.705. The information

Table 2. Conditional item response (column) probabilities, by outcome variable for each class (row) for a model for 3 latent classes

Variable	Pr(1)			Pr(2)		
	Class 1	Class 2	Class 3	Class 1	Class 2	Class 3
A	0.9427	0.4872	0.0000	0.0573	0.5128	1.0000
B	0.8621	0.0000	0.0191	0.1379	1.0000	0.9809
C	1.0000	1.0000	0.1425	0.0000	0.0000	0.8575
D	1.0000	0.9424	0.4138	0.0000	0.0576	0.5862
E	0.9449	0.2494	0.0000	0.0551	0.7506	1.0000
F	1.0000	1.0000	0.5236	0.0000	0.0000	0.4764
G	1.0000	0.3693	0.0000	0.0000	0.6307	1.0000

Source: own calculations in R based on Agresti [2002].

criteria are the following: AIC = 633.41300, BIC = 697.1357, and likelihood ratio *G*2 = 15.2617, and the chi-square goodness of fit statistic χ^2 = 20.5034.

Figure 2 presents the probability of latent class membership for a model with 3 latent variables.

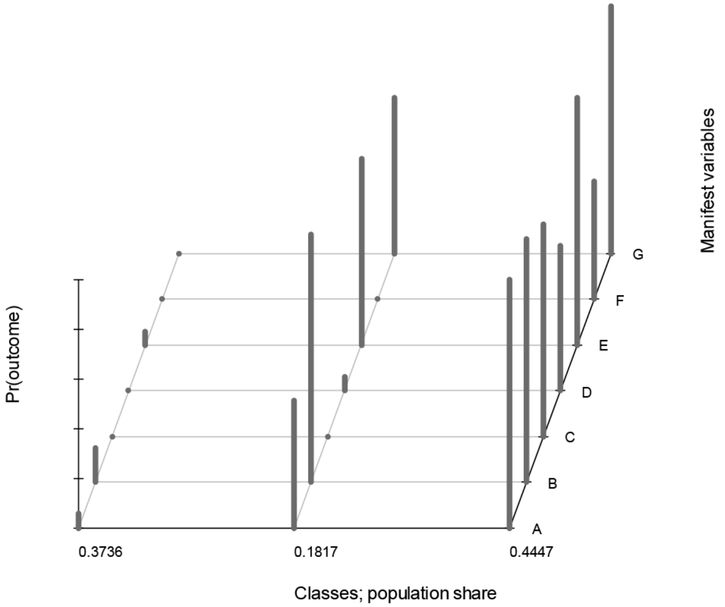

Fig. 2. Probability of latent class membership for a model for 3 latent variables

Source: own calculations in R.

In the next part of this paper we build a model with 4 latent classes. The conditional item response (column) probabilities, by outcome variable for each class (row) are presented in Table 3.

Table 3. Conditional item response (column) probabilities, by outcome variable for each class (row) for a model for 4 latent classes

Variable	Pr(1)				Pr(2)			
	Class 1	Class 2	Class 3	Class 4	Class 1	Class 2	Class 3	Class 4
A	0.0005	0.0000	1.0000	0.4975	0.9995	1.0000	0.0000	0.5025
B	0.6042	0.0191	0.8610	0.0000	0.3958	0.9809	0.1390	1.0000
C	1.0000	0.1412	1.0000	1.0000	0.0000	0.8588	0.0000	0.0000
D	1.0000	0.4133	1.0000	0.9385	0.0000	0.5867	0.0000	0.0615
E	1.0000	0.0000	0.9407	0.2128	0.0000	1.0000	0.0593	0.7872
F	1.0000	0.5229	1.0000	1.0000	0.0000	0.4771	0.0000	0.0000
G	1.0000	0.0000	1.0000	0.3358	0.0000	1.0000	0.0000	0.6642

Source: own calculations in R based on Agresti [2002].

The estimated class population shares are: 0.0281, 0.4441, 0.3543, and 0.1735. The maximum log-likelihood for the model with 3 latent classes is –293.32.

The statistical power to detect the correct number of latent classes in a latent class or latent profile model depends heavily on the selection method or criterion used to determine the number of classes. In this paper the most popular ones were used: AIC, BIC, likelihood ratio and the chi-square goodness of fit statistic. As an alternative, Lo, Mendell, and Rubin (2001), proposed an approximation to the LRT distribution which can be used for comparing nested latent class models. We can also use likelihood-based indexes for model selection or the bootstrap likelihood ratio test (BLRT).

The information criteria are the following: AIC = 648.6399, BIC = 734.5311, and likelihood ratio G^2 = 14.4917, and the chi-square goodness of fit statistic $\chi^2 = 20.8366$.

Figure 3 presents the probability of latent class membership for a model with 4 latent variables.

Comparing the results obtained from the latent class analysis for the model with 2, 3 and 4 latent variables we find that when using information criteria (AIC and BIC) the best model is that with 3 latent classes (AIC = 633.41300, BIC = 697.1357), as the minimal value of those criteria indicate the best fitting model. Because class membership probabilities are modeled as functions of the covariates, and individuals vary with respect to their covariates, there is a vector of estimated class membership probabilities corresponding to each individual (or group of individuals with the same

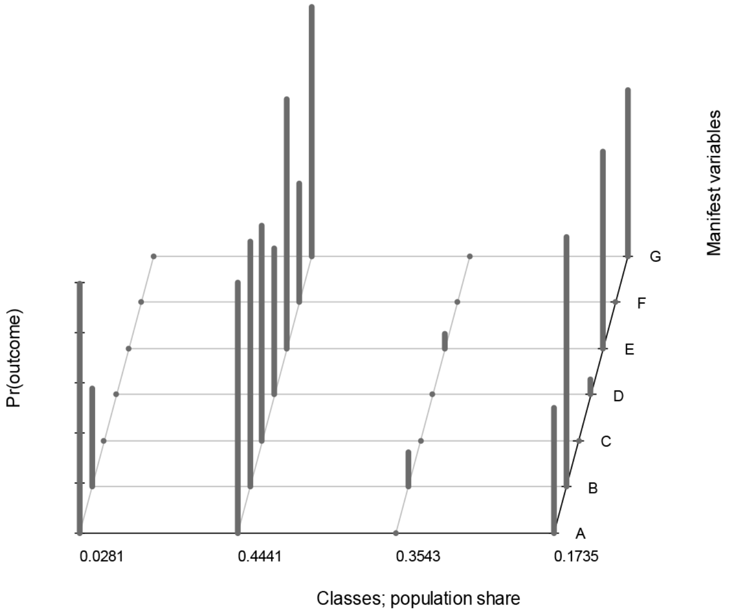

Fig. 3. Probability of latent class membership for a model for 4 latent variables

Source: own calculations in R.

responses to the covariates). The prevalence of each latent class is calculated as the average across participant-specific class membership probabilities. For a model with three latent classes the estimated class population shares are: 0.3736, 0.1817, and 0.4447 for each class respectively.

5. Conclusions

Latent class analysis is a statistical technique for the analysis of multivariate categorical data. This method is an excellent tool for exploring the validity of theoretical notations and hypotheses concerning relations among categorical characteristics. This method can be applied to models with at least one latent variable, as well as covariate and predictor variables. All the calculations were conducted in R software using the `poLCA` package.

In this paper we have shown that it is possible to use discrete multivariate analysis as an effective tool for describing classes. We tested models with 2, 3 and 4 latent classes using the dataset `carcinoma` in the `poLCA` package. We computed information criteria and chi-square based coefficients, as well as presented graphically the probability of latent class membership for each tested model. As a results of the presented analysis using latent class models, we can conclude that the method is an effective and adequate tool for multivariate analysis of cross-classified data providing information on latent class membership. The method can be successfully applied in economic, medical, social and business research.

Bibliography

Agresti A., 2002, *Categorical Data Analysis, second edition*, John Wiley & Sons, Hoboken.

Andersen E.B., 1982, *Latent structure analysis: a survey*, Scand. J. Statist. 9, pp. 1-12.

Anderson T.W., 1954, *On estimation of parameters in latent structure analysis*, Psychometrika, 19, pp. 1-10.

Carroll R., Ruppert D., Stefanski L., Crainiceanu C., 2006, *Measurement error in nonlinear models: a modern perspective*, vol. 105, CRC Monographs on Statistics & Applied Probability, Chapman & Hall, Boca Raton, FL.

Chung H., Anthony J.C., Schafer J.L., 2011, *Latent class profile analysis: an application to stage sequential processes in early onset drinking behaviours*, Journal of the Royal Statistical Society: Series A, 174, s. 689-712.

Clogg C.C., 1981, *Latent structure models of mobility*, American Journal of Sociology, 86, pp. 836-868.

Clogg C.C., Sawyer D.O., 1981, *A Comparison of Alternative Models for Analyzing the Scalability of Response Patterns*, [in:] S. Leinhardt (ed.), *Sociological Methodology 1981*, Jossey-Bass, San Francisco, s. 240-280.

Coull B.A., Agresti A., 1999, *The use of mixed logit models to reflect heterogeneity in capture-recapture* studies, Biometrics, 55, pp. 294-301.

Dempster A.P., Laird N.M., Rubin D.B., 1977, *Maximum likelihood from incomplete data via the EM algorithm*, J. Royal Statist. Soc. B, 39, pp. 1-38.

Edwards J.R., Bagozzi R.P., 2000, *On the nature and direction of relationships between constructs and measures*, Psychol. Methods, 5, pp. 155-74.

Gibson W.A., 1955, *An extension of Anderson's solution for the latent structure equations*, Psychometrika, 20, pp. 69-73.

Green B.F., 1951, *A general solution for the latent class model of latent structure analysis*, Psychometrika, 16, pp. 151-166.

Haberman S.J., 1974, *Log-linear models for frequency tables derived by indirect observation: maximum likelihood equations*, Annals of Statistics, 2, pp. 911-924.

Haberman S.J., 1976, *Generalized residuals for log-linear models,* Proceedings of the ninth International Biometrics Conference, vol. 1 Biometrics Society, Raleigh, NC, pp. 104-172.

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

Hägglund G., 2001, *Milestones in the History of Factor Analysis*, [in:] *Structural Equation Modeling: Present and Future*, ed. R. Cudeck, S. du Toit, D. Sörbom, IL: Scientific Software Int, Lincolnwood.

Hougaard P., 2000, *Analysis of Multivariate Survival Data*, Springer, New York.

Koopmans T.C., 1951, *An Analysis of Production as Efficient Combination of Activities*, [in:] *Activity Analysis of Production and Allocation*, T.C. Koopmans (ed.), Cowles Commission for Research in Economics, Monograph no. 13, New York.

Lazarsfeld P.F., Dudman J., 1951, *The General Solution of the Latent Class Case*, [in:] P.F. Lazarsfeld (ed.), *The Use of Mathematical Models in the Measurement of Attitudes*, RAND Corporation, Santa Monica.

Lo Y., Mendell N., Rubin D., 2001, *Testing the number of components in a normal mixture*, Biometrika, 88, pp. 767-778.

Madansky A., 1959, *The fitting of straight lines when both variables are subject to error*, Journal of the American Statistical Association, 54, pp. 173-205.

Madansky A., 1960, *Determinantial methods in latent class analysis*, Psychometrica, 25, pp. 183-198.

McDonald R.P., 1999, *Test Theory: A Unified Treatment*, Erlbaum, Mahwah, NJ.

McHugh R.B.,1956, *Efficient estimation and local identification in latent class analysis*, Psychometrika, 21, pp. 331-347.

Neale M.C., Cardon L.R., 1992, *Methodology for Genetic Studies of Twins and Families*, Kluwer Academic Publishers, Dordrecht.

Rue H., Held L., 2005, *Gaussian Markov Random Fields: Theory and Applications*, vol. 104 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London.

Shen J., 2009, *Latent class model or mixed logit model? A comparison by transport mode choice data*, Applied Economics, 41, pp. 2915-2924.

Sobel M.E., 1994, *Causal inference in latent variable models*, See von Eye and Clogg, pp. 3-35.

Sutton S., 2000, *Interpreting cross-sectional data on stages of change*, Psychology & Health, 15, pp. 163-171.

Train K.E., 2003, *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

Verbeke G., Molenberghs G., 2000, *Linear Mixed Models for Longitudinal Data*, Springer.

Vermunt J., 1999, *On the use of order-restricted latent class models for defining and testing non-parametric IRT models*, Methods of Psychological Research, 4, 71.

Vermunt J.K., 1997, *Log-linear models for event histories*, Advanced Quantitative Techniques in the Social Sciences Series, vol. 8, Sage Publication, Thousand Oaks.

Wedel M., Kamakura W.A., 2000, *Market segmentation: Conceptual and Methodological Foundations* (second Edition), Dordrecht Kluwer.