Latent Class Modeling and Typological Analysis

Huib Ernste and Manfred M. Fischer
Latent Class Modeling and Typological Analysis

Huib Ernste and Manfred M. Fischer

WSG-Discussion Paper 11

May 1991

ISBN 3 85037 011 9
1. Introduction

In contrast to natural sciences, the measurement of many phenomena in social sciences is discrete rather than metric in nature. This is a consequence of the fact that measurement procedures such as interviews frequently used in social sciences are often confined to measuring on a discrete level. Discrete variables take on values in a limited set of categories which may be on an ordinal or on a purely nominal scale. Recent methodological advances in exploratory and explanatory discrete data analysis provide current researchers with a variety of powerful and flexible approaches for analysing causal relationships between ordinal and nominal dependent and independent variables (see, e.g. Fischer and Nijkamp 1985).

Only very recently latent class analysis is receiving increasing interest in the social sciences (see Arminger 1985, Bartholomew 1987, McCutcheon 1987, Hagenaars 1988, etc.). Latent class analysis essentially deals with the relations between discrete variables. Latent class models contain two kinds of variables: latent (i.e. not directly observable) and manifest (i.e. directly observed) variables. They may be used to identify the latent structures among a set of manifest variables (exploratory mode of analysis) and to test hypotheses about the latent structures among a set of manifest variables (confirmatory mode). In exploratory latent class analysis, the latent class models are unrestricted while in confirmatory latent class analysis they are restricted since a priori size restrictions are imposed on either the conditional probabilities, the latent class probabilities, or both, depending on the hypotheses being tested. Recent developments in latent class analysis enable to analyse the same observed variables to compare the latent variables identified in multiple populations (see Clogg and Goodman 1984, 1985, McCutcheon 1987). The purpose of this paper is to discuss the basic idea of latent class analysis (see section 2) and to illustrate its potentialities for typological analysis through an empirical example, (see section 3).

2. Latent class models: Some fundamental features

Latent class analysis aims at the reduction of many manifest discrete variables to one or few latent discrete variables. Thus, latent class models may be considered as a discrete analogue of factor analysis. There is a great variety of latent class models. In principle, two broad categories of latent class models may be distinguished, viz. unrestricted models for explorative latent class analysis and restricted models for confirmatory latent class analysis. We will first consider the unrestricted case.
Unrestricted models

Assume without loss of generality a general unrestricted latent class model with three discrete manifest variables and a discrete latent variable. Let \( A = (A_i) \) with \( i = 1, \ldots, I \); \( B = (B_j) \) with \( j = 1, \ldots, J \) and \( C = (C_k) \) with \( k = 1, \ldots, K \) denote three (dichotomous or polytomous) discrete manifest variables. These three variables constitute a three-way contingency table which cross-classifies a sample of \( n \) individuals with respect to the observed variables \( A, B \) and \( C \). Suppose that there is a (polytomous) discrete latent variable \( X \) with \( T \) categories (termed ‘latent classes’) which is able to explain the relationships among the variables in the three-way contingency table. Assume also that each individual in the population concerned belongs to one and only one of the \( T \) latent classes. Then

\[
\pi_{ijk}^{ABC} = \sum_{t=1}^{T} \pi_{ijkt}^{ABCX}
\]

where \( \pi_{ijk}^{ABC} = p(A = i, B = j, C = k) \) is the probability of an event being in cell \((i,j,k)\) of the observed three-way contingency table and \( \pi_{ijkt}^{ABCX} \) the (unknown) probability of an observation belonging to cell \((i,j,k,t)\) of the four-way table cross-classifying the observed variables \( A, B \) and \( C \) and the unobserved latent variable \( X \).

Essential to the standard latent class model is the assumption of local independence, i.e. that the latent variable completely accounts for the manifest variables. This assumption implies that the three manifest variables are all conditionally independent of each other, given level \( t \) of \( X \). Thus

\[
\pi_{ijkt}^{ABCX} = \pi_{it}^{(A \mid X)} \pi_{jt}^{(B \mid X)} \pi_{kt}^{(C \mid X)} \pi_{t}^{X}
\]

where

\[
\pi_{t}^{X} = p(X = t) = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \pi_{ijkt}^{ABCX}
\]

is the probability of a case being located in latent class \( t \) (the so-called latent class probability) and \( \pi_{it}^{(A \mid X)} \) the conditional probability of category \( i \) of \( A \), given the value \( t \) of \( X \). \( \pi_{jt}^{(B \mid X)} \) and \( \pi_{kt}^{(C \mid X)} \) are defined analogously. Because the parameters in (2) are
probabilities they are subject to the following constraints:

\[ \sum_{t=1}^{T} \pi_i^X = \sum_{i=1}^{I} \pi_{it}^{(A \mid X)} = \sum_{j=1}^{J} \pi_{jt}^{(B \mid X)} = \sum_{k=1}^{K} \pi_{kt}^{(C \mid X)} = 1 \]  

\((\pi_i^X, \pi_{it}^{(A \mid X)}, \pi_{jt}^{(B \mid X)}, \pi_{kt}^{(C \mid X)})\) denotes the vector of parameters to be estimated in the general latent class model. The latent class probabilities \(\pi_i^X\) describe the distribution of classes of the latent variable \(X\). There are two important aspects of the latent class probabilities: the number of classes \(T\) representing the number of latent classes defined by the model and the relative sizes of these classes indicating whether the population is relatively evenly distributed among the \(T\) classes or not. The second type of parameters, the conditional probabilities - comparable to the factor loading in factor analysis - represent the probabilities of an individual in class \(t\) of the latent variable \(X\) being at a particular level of the manifest variables (see McCutcheon 1987).

Standard methods can be used to prove that the maximum likelihood estimates of (2)-(4), denoted by circumflexes of the corresponding parameters, satisfy the following set of equations (see Goodman 1979, Haberman 1979):

\[ \hat{\pi}_i^X = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} p_{ijk} \hat{\pi}_{ijk} \pi_{i}^{(X \mid ABC)} \]  

\[ \hat{\pi}_{it}^{(A \mid X)} = \sum_{j=1}^{J} \sum_{k=1}^{K} p_{ijk} \hat{\pi}_{ijk} \pi_{i}^{(X \mid ABC)} \]  

\[ \hat{\pi}_{jt}^{(B \mid X)} = \sum_{i=1}^{I} \sum_{k=1}^{K} p_{ijk} \hat{\pi}_{ijk} \pi_{i}^{(X \mid ABC)} \]  

\[ \hat{\pi}_{kt}^{(C \mid X)} = \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ijk} \hat{\pi}_{ijk} \pi_{i}^{(X \mid ABC)} \]
where \( p_{ijk}^{ABC} \) is the observed proportion of individuals at the level \((i, j, k)\) of a cross tabulation of manifest variables \((A, B, C)\). Note that the conditional probabilities \( (X | ABC) \) can be expressed in terms of \( \pi_{ijk} \) and \( \pi_{ijkt} \) as follows

\[
\frac{(X | ABC)}{\pi_{ijk}} = \frac{\pi_{ijkt}^{ABC}}{\pi_{ijk}^{ABCX}}
\]

with

\[
\pi_{ijk}^{ABC} = \sum_{t=1}^{T} \pi_{ijkt}^{ABCX}
\]

Because of condition (10) only \( T - 1 \) of the \( \pi_i^X \), \( 1 - 1 \) of the \( \pi_{it}^{(A | X)} \), \( J - 1 \) of the \( \pi_{jt}^{(B | X)} \) and \( K - 1 \) of the \( \pi_{kt}^{(C | X)} \) have to be estimated. The ML-estimates of the parameters may be obtained through the iterative proportional fitting procedure. Fisher's scoring procedure, a variant of the Newton-Raphson procedure, may also be applied to any latent class model based on a log-linear model. In this latter case, latent class modelling can be considered as a special case of general linear models (GLM) with a composite link function defined by Thomson and Baker (1981) (see also Arminger 1985). More details on estimation problems can be found in Goodman (1978, p.542). Despite a general resemblance of the maximum likelihood-equations of this general latent class model to those for ordinary log-linear models, there are many difficulties in contrast to ordinary log-linear modelling, especially because maximum likelihood-estimates are not uniquely determined (leading to multiple solutions or to solutions which are not ML-estimates).

Thus, special attention must be paid to identification problems of a ML-estimate (cf. Haberman 1979, p.542). A necessary and sufficient condition for determining the local identifiability of a latent class model is that the \(((I + J + K - 2) T - 1)\) matrix of partial derivatives of the non-redundant observed probabilities with respect to the non-redundant model parameters is of full column rank, equal to \((I + J + K - 2) T - 1\).

The fit of model to data can be assessed in two ways. The first is the Pearson \( \chi^2 \) statistic.
\[
\chi^2 = n \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left( p_{ijk}^{ABC} - \hat{\pi}_{ijk}^{ABC} \right)^2 / \hat{\pi}_{ijk}
\]

(11)

and the second is the \( \chi^2 \) statistic based on the likelihood-ratio criteria

\[
L^2 = 2n \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} p_{ijk}^{ABC} \log \left( \frac{p_{ijk}^{ABC}}{\hat{\pi}_{ijk}} \right)
\]

(12)

where \( p_{ijk}^{ABC} \) denotes the observed proportions. Both statistics are justified by large sample \( \chi^2 \) distribution because it is a quadratic form and (12) by virtue of Wilks’ theorem which states that the logarithm of the ratio of two likelihood functions is proportional to a \( \chi^2 \) random variate. For very large samples it has been shown that \( \chi^2 \) is superior to \( L^2 \).

3. Searching for Types: An Empirical Example

As noted in the previous section latent class analysis permits to characterize nominal level latent variables from a cross-tabulation of two or more nominal or ordinal manifest variables. The potentialities of latent class analysis will be illustrated by data from the 1980 Swiss population census. The basic data file for the city of Zürich consists of 380,529 individuals.

For the purpose of illustration let us consider a latent classification of types of permanent residents where permanent residents are defined as those individuals who lived in Zürich in 1975 and did not move in the time period of 1975 to 1980. A sample of 1,114 individuals out of the total population of 275,060 permanent residents was drawn randomly. Five categorically scored variables have been chosen for the analysis: sex (male/female), household size (singles / multi-person households), nationality (Swiss / foreign), age (0-15 years / 16-25 years / 26-40 years / 41-64 years / 65 and more years), and employment and vocational status (non-active / part-time / full time and low vocational status / full time and lower middle vocational status / full time and higher middle vocational status / full time and high vocational status).

These five variables net a \((2 \times 2 \times 2 \times 5 \times 5)\) 200 cell cross-tabulation. For notational convenience the five variables are designated as A, B, C, D and E, indexed by i, j, k, l, m respectively. The LCAG program package developed by Hagenaars et al. (1990) is used to carry out the study.
Like (exploratory) factor analysis, (exploratory) latent class analysis is a useful approach for data reduction and enables to characterise the structure of the latent topology by analysing the conditional probabilities for each of the observed variables in each of the latent classes.

Table 1: Unrestricted Latent Class Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Degrees of Freedom</th>
<th>Goodness-of-Fit Chi-Square $\chi^2$</th>
<th>Likelihood-Ratio-Chi-Square $L^2$</th>
<th>Probability Value of $L^2$</th>
<th>Decision at 0.05 Significance Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete Independence (One-Class) Model</td>
<td>227</td>
<td>1513.47</td>
<td>1189.52</td>
<td>&lt;0.0000</td>
<td>reject</td>
</tr>
<tr>
<td>Two-Class Model</td>
<td>219</td>
<td>684.36</td>
<td>616.42</td>
<td>&lt;0.0000</td>
<td>reject</td>
</tr>
<tr>
<td>Three-Class Model</td>
<td>211</td>
<td>500.51</td>
<td>534.37</td>
<td>&lt;0.0000</td>
<td>reject</td>
</tr>
<tr>
<td>Four-Class Model</td>
<td>205</td>
<td>383.36</td>
<td>390.99</td>
<td>&lt;0.0000</td>
<td>reject</td>
</tr>
<tr>
<td>Five-Class Model</td>
<td>198</td>
<td>333.03</td>
<td>253.24</td>
<td>0.0049</td>
<td>reject</td>
</tr>
<tr>
<td>Six-Class Model</td>
<td>193</td>
<td>182.01</td>
<td>169.78</td>
<td>0.885</td>
<td>accept</td>
</tr>
</tbody>
</table>

In applying the standard latent class model to the multidimensional table formed by the five variables, one first has to decide upon the number of categories of the latent variable, that is, the number of latent classes. This decision can be made on theoretical grounds if one has a clear notion about the number of types. Otherwise, in a more exploratory manner one tries several latent class models with a different number of latent classes, and chooses the model that fits the observed data and lends itself best to a meaningful interpretation.

The decisions about the acceptability of the fit among several latent class models and the observed data are reported in table 1. The likelihood ratio chi-square rather than the Pearson chi-square is the preferred test statistic because it is possible to partition $L^2$. This property means that once the number of classes (T) in the latent variable has been determined hypotheses about the values of the conditional probabilities and latent class probabilities can be efficiently tested. This is especially important for the confirmatory
mode of latent class analysis which will be discussed later. It is also worthwhile to mention that the decision criterion is set at the $p > 0.05$ alpha level.

The first model in Table 1, the complete independence model, is equivalent to test a latent class model with only a single latent class, i.e. $T=1$, within which all of the observed variables are locally independent. If this model would be acceptable, the manifest variables would be not interrelated and consequently no latent variable would be needed.

Table 1, however, shows that the complete independence model for the data at hand has to be rejected with $L^2 = 1189.52$ and $\chi^2 = 1513.47$ (227 d.f.). The next hypothesis tests that there are two latent population classes. Although the two-class model provides a much better fit ($L^2 = 616.42$ and $\chi^2 = 684.36$ with 219 d.f.) than the complete independence model, the model has to be rejected at the 0.05 alpha level. Finally, the six-class model provides an acceptable fit of the observed data ($L^2 = 169.78$ and $\chi^2 = 182.01$ with 193 d.f.). The exploratory analysis leads to six latent classes or types.

In Table 2 the conditional and latent class probabilities are summarized. The last row presents the latent class probabilities which indicate that the individuals are unevenly divided among the six latent classes. The fourth latent class contains about one third, the sixth latent class about one fourth and the other types each about 10 per cent of all cases.

The remaining cell entries in Table 2 refer to the conditional probability ($\pi_{it}^{A/X}$) which indicate that an individual obtains a particular score (i) on a manifest variable (A), given the particular latent class (t) he or she belongs to. For example, the cells at the upper left corner show that an individual who belongs to the first latent class is male then of course equals 0.92.

What are the characteristics of the six latent classes of permanent residents identified in the analysis? The first class corresponds most closely to females (0.92). About 9 of 10 are Swiss, seven of ten 16 to 40 years old, nine of ten work full time. The variables distinguishing the second from the first type most markedly are household size and employment status, to a lesser extent age too. The second class may be characterised as consisting of middle aged females who are members of a two or more person household and work at home or part time. The third class is the smallest one in the typology. The variable distinguishing this class from the others is nationality. All individuals belonging to the class are foreign and employed full-time, typically with a lower vocational status.
About three fourth are male. The fourth class - the largest of the typology - corresponds closely to elderly Swiss (0.95) in the retirement stage (0.85). Three of four are female. The fifth type consists of Swiss kids, teenagers and twens living with their parents and going to school. The variable distinguishing this class most markedly from all the others is age and to a lesser degree employment status, too. The final latent class might perhaps best be labeled as Swiss family fathers with a higher vocational status.

Table 2: Conditional and Latent Class Probabilities for the Six Class Model

<table>
<thead>
<tr>
<th>Manifest Variables</th>
<th>Categories</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Male</td>
<td>0.08</td>
<td>0.00</td>
<td>0.72</td>
<td>0.25</td>
<td>0.44</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>0.92</td>
<td>1.00</td>
<td>0.28</td>
<td>0.75</td>
<td>0.56</td>
<td>0.00</td>
</tr>
<tr>
<td>Household</td>
<td>One Person</td>
<td>0.45</td>
<td>0.00</td>
<td>0.24</td>
<td>0.34</td>
<td>0.00</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Two and more Persons</td>
<td>0.55</td>
<td>1.00</td>
<td>0.76</td>
<td>0.66</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td>Nationality</td>
<td>Swiss</td>
<td>0.94</td>
<td>0.84</td>
<td>0.00</td>
<td>0.95</td>
<td>0.82</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Foreign</td>
<td>0.06</td>
<td>0.16</td>
<td>1.00</td>
<td>0.05</td>
<td>0.18</td>
<td>0.00</td>
</tr>
<tr>
<td>Age</td>
<td>0-15 Years</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.81</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>16-25 Years</td>
<td>0.29</td>
<td>0.00*</td>
<td>0.11</td>
<td>0.05</td>
<td>0.15</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>26-40 Years</td>
<td>0.43</td>
<td>0.42</td>
<td>0.47</td>
<td>0.00</td>
<td>0.03</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>41-64 Years</td>
<td>0.12</td>
<td>0.58</td>
<td>0.29</td>
<td>0.00</td>
<td>0.00</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>65 Years and more</td>
<td>0.16</td>
<td>0.00</td>
<td>0.12</td>
<td>0.95</td>
<td>0.00</td>
<td>0.27</td>
</tr>
<tr>
<td>Employment and</td>
<td>Non-Active</td>
<td>0.00</td>
<td>0.60</td>
<td>0.00</td>
<td>0.85</td>
<td>1.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Vocational Status</td>
<td>Part-Time</td>
<td>0.07</td>
<td>0.38</td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Full-Time, Low</td>
<td>0.02</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>Full-Time, Lower Middle</td>
<td>0.28</td>
<td>0.00</td>
<td>0.44</td>
<td>0.00</td>
<td>0.00</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Full-Time, Higher Middle</td>
<td>0.63</td>
<td>0.00</td>
<td>0.26</td>
<td>0.04</td>
<td>0.00</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Full-Time, High</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>Latent Class Probabilities</td>
<td>0.13</td>
<td>0.10</td>
<td>0.08</td>
<td>0.34</td>
<td>0.11</td>
<td>0.24</td>
<td></td>
</tr>
</tbody>
</table>

* zero by rounding (0.0017)
Frequently, exploratory latent class analysis is only a first step in research. Usually further analyses are carried out by assigning individuals to the latent classes on the basis of their scores on the manifest variables. But as in common factor analysis, two kinds of problems may arise in respect. First, misclassifications may occur, and a second problem might arise that within the latent class model the individual scores on the latent variable are not exactly identified (see Hagenaars and Habman 1989, McCutcheon 1987 for more details).

The exploratory latent class analysis imposes no restrictions on the values that the conditional and/or latent class probabilities could take. In general, there are two types of restrictions which can be placed on each of the two types of parameters: equality constraints and specific value constraints. Equality constraints require that two or more of the conditional or latent class probabilities take on the same value. In the case of equal conditional probabilities hypotheses regarding the nature of the contributions of the manifest variables to the latent classes are tested, while the hypothesis that classes are of equal size may be tested via equal latent class possibilities.

For example, in the six class model one might be interested in testing the hypotheses that the variable age or certain age classes do not discriminate between latent class 1 and latent class 3. One might state that the probability of being in the age class 16-25 years, given an individual is in latent class 1 (X=1), is the same as the probability of being in this age class, given an individual is in latent class 3 (X=3):

\[ \pi('age'/X)_{16-25,1} = \pi('age'/X)_{16-25,3} \] (13)

In an analogous manner similar equality restrictions might be imposed concerning the age classes 26-40 years, 41-64 years and 65 years and older:

\[ \pi('age'/X)_{26-40,1} = \pi('age'/X)_{26-40,3} \] (14)

\[ \pi('age'/X)_{41-64,1} = \pi('age'/X)_{41-64,3} \] (15)

\[ \pi('age'/X)_{\geq 65,1} = \pi('age'/X)_{\geq 65,3} \] (16)

As shown in table 3, the addition of restrictions increases in the L2-value of the six class model. Since the unrestricted six class model and the restricted models with one and two constraints are nested, L2 can be partitioned to check for the acceptability of the model restrictions. For example, to test the acceptability of restriction (13), the L2-value (and
degrees of freedom) of the restricted model (169.78) has to be subtracted from the $L^2$-value (and degrees of freedom) of the restricted model (177.96) which results in the unacceptable increase in $L^2$ (8.18 with 194-193 = 1 d.f.). The increase of $L^2$ nets a probability that is less than 0.05. Thus, the equality restriction causes the model to differ significantly from the data. Hypothesis (13) has to be rejected or in other words age class 16-25 years discriminates between the two latent classes considered. In contrast to equality restriction (13), constraint (14) leads to a $L^2$ increase of only 0.33 which nets a probability of 0.567. Consequently, it might concluded that this restriction - and the same is true for restriction (16) - substantially improves the fit of the model.

Table 3: Confirmatory Latent Class Models: Equality Restrictions

<table>
<thead>
<tr>
<th>Model</th>
<th>Degrees of Freedom</th>
<th>Goodness-of-Fit Chi-Square $\chi^2$</th>
<th>Likelihood-Ratio Chi-Square $L^2$</th>
<th>Probability Value of $L^2$ ($\Delta L^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted</td>
<td>193</td>
<td>182.01</td>
<td>169.78</td>
<td>0.885</td>
</tr>
<tr>
<td>Six-Class Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equality Restriction (13)</td>
<td>194</td>
<td>197.22</td>
<td>177.96</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Equality Restriction (14)</td>
<td>194</td>
<td>181.22</td>
<td>170.10</td>
<td>(0.567)</td>
</tr>
<tr>
<td>Equality Restriction (15)</td>
<td>194</td>
<td>193.22</td>
<td>177.61</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Equality Restriction (16)</td>
<td>194</td>
<td>181.20</td>
<td>169.94</td>
<td>(0.689)</td>
</tr>
<tr>
<td>Equality Restrictions (13) - (16)</td>
<td>196</td>
<td>202.16</td>
<td>183.82</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>

Another useful feature of confirmatory latent class analysis is the possibility to assign fixed values to the parameters of the latent class model, especially the extreme values zero or one. Sometimes a manifest variable A can be conceived of as a perfect indicator of the underlying typology X. This kind of relationship between A and X can be taken into account by imposing the following constraints $\pi_{it}^{(A1X)} = 1$ if $i=t$, $\pi_{it}^{(A1X)} = 0$ otherwise. The manifest variable A and the latent variable X are then identical variables, and thus A often will be termed a quasi-latent variable. Quasi-latent variables may be used for several purposes. For example, if one aims to compare the typology established for a specific cultural context at one point in time with typologies found for other cultural contexts or other time periods. Such cross-cultural or over-time comparisons can be made by means of a latent class model in which the cultural context variable or time variable has been
included as a quasi-latent variable (for more details see Clogg and Goodman 1985, Hagenaars 1988).

4. Conclusions and Outlook

Latent class analysis provides a flexible methodological approach for discovering and testing the "existence of ideal types" in the sense of Max Weber. Latent class modelling does not assume interval measurements, linear relationships, or underlying normal distributions usually done in alternative approaches such as factor analysis. Most statistical problems with latent class analysis - such as how and when to obtain estimates of identifiable parameters - have been overcome, especially through the work of Haberman and Goodman (see, for example Haberman 1979, 1989, Goodman 1978, 1979). Computer programs such as Haberman's LAT, Clogg's MLLSA and Hagenaar's LCAG are available to routinely carry out the analyses (Haberman 1979, Clogg 1981, Hagenaars 1988). Nevertheless, one major problem remains to be unsolved, a problem arising with all forms of table analysis. If the number of manifest variables increases, the number of cells increases enormously, and may exceed the number of individuals. Under such conditions the standard chi-square testing procedures are not useful. Only, recently, there are some attempts to tackle this problem (see, for example, Koehler 1986).

Apart from modifications of the standard model discussed in this contribution most new developments focus on the formulation of latent class model as a log-linear model with latent variable (see, for example, Hagenaars 1988).
References


