Appendix A.

Basics of Latent Class Analysis

Latent class analysis (LCA) is a statistical method for discovering subtypes of related cases from multivariate categorical data. A latent class may be considered as an abstract characteristic or trait that cannot be observed directly. If an abstract latent construct is believed to be of a discrete nature, that is, to have qualitatively different characteristics or traits within it, then each characteristic or trait is considered as a latent class. Most traits studied in the social and behavioral sciences can be considered as characterizations of latent constructs. For example, social development in countries or religious commitment in people can be viewed as characterizing latent classes.

Originally developed to assist in survey data analysis, applications of LCA now range from identifying disease subtypes or political belief to marketing research uses. Applications of LCA to psychiatric epidemiologic research have been relatively new (Eaton, McCutcheon, Dryman, & Sorenson, 1989), but its application is increasing rapidly, in part because of its ability to derive discrete classes. It is suitable for analyses based on the categorical classification that are typical in psychiatry (Eaves, et al., 1993; Hudziak, et al., 1998).

The basic premise of LCA models is that the covariation actually observed among the observed variables is due to each observed variable’s relationship to the latent variable (McCutcheon, 1987). Latent classes are defined by the criterion of “conditional independence,” meaning that, within each latent class, each variable is statistically independent of every other variable. LCA (Goodman, 1978; McCutcheon, 1987) enables estimations of the probability that any given individual would fall into
a specific latent class. Given the class that the individual is in, LCA also estimates the probability that the individual would endorse a certain behavior or trait.

For the current monograph, the latent classes comprise a behavioral construct, such as childhood conduct syndrome, and the observed variables are the individual symptoms, such as truancy and school expulsion. To describe the basics of LCA, an example is derived from the estimation for the adult manifestation of antisocial personality disorder among males in the United States (St. Louis). The observed variables comprised seven of the nine symptoms of adult antisocial personality disorder: work problems, illegal activities, marital or relationship problems, violence, vagrancy, lying, and traffic offences (See Table 6.1.).

The class probabilities and the endorsement rates for each class are the unknown variables of an LCA model. The individuals can be divided into \( n \) classes; in the example of adult antisocial personality above, \( n = 2 \) was eventually chosen. The probability of an individual being in class \( n \) is \( L_n \), since every observation must be in a class: \( L_1 + L_2 = 1 \).

For any observed measure, such as work problems as a symptom of adult ASP, the probability of giving a specific answer is determined by the individual’s latent class membership. The probability of endorsing work problems \( (W_{n1}) \) would be the same for all respondents in class \( n \). The same would be true for the probability of not endorsing work problems \( (W_{n2}) \). Since endorsing and not endorsing are the only two options in this case, \( W_{n1} + W_{n2} = 1 \). These also hold for other six measures: illegal activities \( (A_{nj}) \), marital or relationship problems \( (M_{nj}) \), violence \( (V_{nj}) \), vagrancy \( (G_{nj}) \), lying \( (L_{nj}) \), and traffic offenses \( (T_{nj}) \).

The computer applications such as MLLSA (MLLSA, 2000a) or WinLTA (WinLTA, 2000b) then choose the values for the probabilities that make the observed outcomes most likely. In this
example, the probability of observing an outcome would be:

\[ L_1 W_{ij} A_{ijj} M_{ijj} V_{ijj} G_{ijj} L_{ijj} T_{ijj} + L_2 W_{2ij} A_{2ij} M_{2ij} V_{2ijj} G_{2ij} L_{2ij} T_{2ij}. \]  

(A.1)

If \( N \) is the total number of individuals then the probability of observing the entire data set is:

\[ P = D (L_1 W_{ij} A_{ijj} M_{ijj} V_{ijj} G_{ijj} L_{ijj} T_{ijj} + L_2 W_{2ij} A_{2ij} M_{2ij} V_{2ijj} G_{2ij} L_{2ij} T_{2ij}); \]  

(A.2)

where, \( D \) is the product over all \( N \) individuals. MLLSA and WinLTA both select the values of the variables that maximize \( P \) by use of the Expectation-Maximization (EM) algorithm.

In addition to the values that maximize \( P \), the LCA applications also supply measures of how badly the model fits the data. These measures are the likelihood ratio \( L^2 \) and the Pearson chi-square \( \chi^2 \). For simple models, they both should behave as chi-square distributions with the number of degrees of freedom depending on the number of total possible outcomes and the number of free parameters.

In evaluating models, the standard parsimonious method can be used for simple models. In this method, the number of latent classes is increased until the \( L^2 \) score is not rejected at the given number of degrees of freedom. Selection of proper number of classes, however, sometimes involves judgment calls, when the combinations of endorsement items is very large, which could result in a large number of “empty cells” that is, combinations of endorsements that no sample members endorsed. In a complicated model, it is well known that a large number of empty cells greatly inflate the degrees of freedom. In such a case, the nested-model method is used where a model with \( n-1 \) classes is compared with a model with \( n \) classes. This method is equivalent to equating the endorsement rates of two classes to form one larger class in the \( n-1 \) model. Though not without its own methodological problems, the difference of the \( L^2 \) values in the nested model method should be in line with the difference in the degrees of freedom.
In the above example of adult ASP among U.S. men, two latent classes were selected, that luckily satisfied both parsimonious and nested-model methods. Individuals had a 73% chance of falling into the first latent class that was described as “not affected” and a 27% chance of falling into the second latent class that was described as “affected.” These are called class probabilities. The latent class that the observation belongs to determines that probability of endorsing the observed variables. The probabilities for all seven symptoms given the latent class are shown in Table A.1. For individuals in the “not affected” latent class, the probability of endorsing the symptom of work problems was 37%, and for individuals in the “affected” latent class the probability of endorsing work problems was 84%. Legal problems were less commonly reported overall, so that the probability of endorsing legal problems was 2% among those in the “not-affected” latent class, and 35% even in the “affected” latent class.

— Table A.1. About Here —

For ease of comparison, these probabilities, or rates, are often displayed in a graphic format. Each class is represented as line on a graph as in Figure A.1, which is itself a component of Table 6.2 appearing in Chapter 6. All endorsement rates are described in graphic format in Chapter 6 through Chapter 8. Class probabilities are noted in the same graphs throughout.

— Figure A.1. About Here —
Table A.1. Endorsement Rates of Adult Antisocial Personality Disorder for Saint Louis Males

<table>
<thead>
<tr>
<th></th>
<th>“Not Affected”</th>
<th>“Affected”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Problems</td>
<td>.37</td>
<td>.84</td>
</tr>
<tr>
<td>Illegal Activities</td>
<td>.02</td>
<td>.35</td>
</tr>
<tr>
<td>Marital Problems</td>
<td>.12</td>
<td>.51</td>
</tr>
<tr>
<td>Violence</td>
<td>.17</td>
<td>.68</td>
</tr>
<tr>
<td>Vagrancy</td>
<td>.02</td>
<td>.33</td>
</tr>
<tr>
<td>Lying</td>
<td>.02</td>
<td>.22</td>
</tr>
<tr>
<td>Traffic Offenses</td>
<td>.28</td>
<td>.70</td>
</tr>
</tbody>
</table>
Figure A.1 Endorsement Rates of Adult Antisocial Personality Disorder for St. Louis Males

Class 1 - 73%
Class 2 - 27%
Reference List


MLLSA (2000). MLLSA. Available at: www.avalon.net/~seliason. [Computer program].

WinLTA (2000). WinLTA. Available at: mathcenter.psu.edu/WinLTA.html. [Computer program].