

## Latent Class Analysis for Models with Error of Measurement Using Log-Linear Models and An Application to Women's Liberation Data

Haydar Demirhan  
*Hacettepe University*

*Abstract:* This article deals with the latent class analysis of models with error of measurement. If the latent variable is ordinal and manifest variables are nominal, an approach to handle the restrictions is given for latent class analysis of the models with error of measurement using log linear models. By this way, we include ordinal nature of the latent variable into the analysis. Therefore, overall uncertainty is decreased, and our inferences become more precise. The new approach is applied to a women's liberation data set.

*Key words:* Item specific, Lazarsfeld's latent distance, nominal, ordinal, Proctor, true type specific, Guttman scales.

### 1. Introduction

Latent class analysis is frequently used in social sciences and education. Main aim of the analysis is to explain the association structure between manifest variables by using unobserved variables, namely latent variables. Latent class analysis is a categorical analogous of the factor analysis when the latent and manifest variables are categorical. Log-linear models are widely used for the analysis of contingency tables. It is possible to represent a latent class model as a log-linear model using conditional response probabilities. This representation is called as log-linear parametrization. This is a special case of Formann's linear logistic latent class analysis (Formann, 1992).

Error of measurement models are probabilistic versions of Guttman scale (Guttman, 1950), and considered as restricted latent class models. In the log-linear parametrization of the latent class models, types of manifest and latent variables are important issues because latent class models specialize according to the typology of the variables. For example, if the latent variable is metrical and manifest variables are nominal then appropriate analysis is to carry on latent class analysis with linear restrictions or use nominal response models (Heinen, 1996). In our concerned models with error of measurement, manifest variables

can be nominal or ordinal, but latent variable is ordinal. If manifest variables are also ordinal then latent class analysis with ordinal classes is a convenient way. However, if they are nominal while latent variable is ordinal, there is no proposed latent class analysis; and latent variable is treated as a metric variable in this case (Heinen, 1996). This is an inappropriate approach for the models with error of measurement. Type of used log linear model is another issue here. Heinen (1996) uses independence log linear models for the latent class analysis of models with error of measurement. In this, ordinal nature of latent variable is ignored. Also, Heinen (1996) mentions use of logit form of column or row association models when column or row variable of a two-way classification is metric, and use of logit form of uniform association models when latent and manifest variables are metric in a two-way classification for latent class analysis. However, these models should be used for ordinal variables rather than metric variables, and use of them for models with error of measurement is not noted. In this article, we propose to use column or row effects log linear models in the analysis of models with error of measurement over multi-way tables when the latent variable is ordinal and manifest variables are nominal.

We revisit the women's liberation data set of Felling *et al.* (1987). It is also analyzed by Heinen (1996) over models with error of measurement. He treats the ordinal latent variable as if it is metric, and uses an independence log linear model. Our approach is a more convenient option for the analysis of this data set because we take into consideration of ordinal nature of latent variable using an appropriate log linear model. We clarify our approach over the data set, compare our results with those obtained by Heinen (1996), and evaluate effects of both approaches for the models with error of measurement over the data set. By this application, we aim to promote use of column or row effects models in the analysis of latent class models with error of measurement for this kind of data sets.

Section 2 explains the log-linear model and necessary notation. Section 3 gives the basic latent class model. Section 4 explains models with error of measurement. Log-linear approach to the latent class models is given in Section 5 by means of nominal and ordinal latent variable. An expectation-maximization (EM) algorithm for estimation procedure is also presented in Section 5. The women's liberation data is analyzed using the given approach over the models with error of measurement in Section 6.

## 2. The Log-linear Model

Log-linear models vary according to the type of variables that construct the contingency table. Classical independence model is used when all variables are nominal; however, if they are ordinal, interaction models are used. If some variables are nominal and some are ordinal, row or column effect models are used.

Notation given in this section are valid under all cases. Furthermore, it is assumed that effect coding is used. An alternative to effect coding is dummy coding, which sometimes makes the representations easier (Heinen, 1996).

Let  $S_1$ ,  $S_2$  and  $S_3$  be nominal categorical random variables (r.v.'s) constituting a  $R \times C \times K$  contingency table, and  $|\cdot|$  represent number of levels of inner categorical r.v. Hence  $|S_1| = R$ ,  $|S_2| = C$  and  $|S_3| = K$ . The representation of a saturated log-linear model is as follows:

$$\log n_{ijk} = \beta + \beta_i^1 + \beta_j^2 + \beta_k^3 + \beta_{ij}^{12} + \beta_{ik}^{13} + \beta_{jk}^{23} + \beta_{ijk}^{123}, \quad (2.1)$$

where  $i = 1, \dots, R$ ,  $j = 1, \dots, C$ ,  $k = 1, \dots, K$ ;  $\log n_{ijk}$  is the natural logarithm of expected cell count corresponding to  $i$ ,  $j$  and  $k$  levels of the first, second and third variables;  $\beta$  is normalizing constant;  $\beta_i^1$ ,  $\beta_j^2$  and  $\beta_k^3$  are main effect parameters of levels  $i$ ,  $j$  and  $k$  of the first, second and third variables, respectively;  $\beta_{ij}^{12}$ ,  $\beta_{ik}^{13}$  and  $\beta_{jk}^{23}$  are interaction effects of the corresponding levels of the first and second, first and third, and second and third variables, respectively.

Let  $S_1$ ,  $S_2$  be nominal and  $S_3$  be an ordinal categorical r.v. constituting a  $R \times C \times K$  contingency table. Then the log-linear model as a row (column) effect model is represented as follows:

$$\log n_{ijk} = \beta + \beta_i^1 + \beta_j^2 + \beta_k^3 + \beta_{ij}^{12} + \tau_i^{13} \nu_k + \tau_j^{23} \nu_k, \quad (2.2)$$

where  $\tau_i^{13}$  and  $\tau_j^{23}$  are row (column) effect parameters of nominal-ordinal interactions of the first and third, and the second and third variables, respectively;  $\nu_k$  is the score value corresponding to level  $k$  of the ordinal variable; definitions of the rest of the elements of eq. (2.2) are the same as in eq. (2.1). Similar models are constructed according to the number of ordinal variables. See Agresti (2002) for more details.

### 3. Latent Class Models

There are two main variable types in the latent class model: manifest and latent variables. Manifest variables are directly obtained and contain information about latent variable, and latent variables are theoretical and not observed directly. Let  $X$  be the latent variable, using the representations of Section 2,  $|X|$  is the number of unobserved latent classes. When the manifest variables come from  $N$  r.v.'s, there are  $|X| \cdot |S_1| \cdots |S_N|$  cells in the incomplete contingency table. Because contingency table of interest is incomplete, the aim of latent class analysis can be pertained as to complete the table.

Main assumption of the latent class analysis is the local independence. Although the manifest variables constituting the complete contingency table are interrelated with each other, if they are independent on the levels of the latent

variable  $X$ , this association structure is defined as local independence by Lazarsfeld and Henry(1968). For  $N = 3$ ,  $\ell = 1, \dots, |X|$ , basic unrestricted latent class model is as follows:

$$\pi_{ijk\ell}^{S_1 S_2 S_3 X} = \pi_\ell^X \prod_{\lambda=1}^N \pi_{g\ell}^{\overline{S_\lambda} X} \quad (3.1)$$

where  $g$  equals to  $i$ ,  $j$  or  $k$  if  $\lambda$  equals to 1, 2 or 3, respectively;  $\pi_{ijk\ell}^{S_1 S_2 S_3 X}$  is the probability of being in the cell  $ijk\ell$  of the incomplete table for a randomly selected individual;  $\pi_\ell^X$  is the probability that a randomly selected individual is in the level  $\ell$  of the latent variable, which are called latent class probabilities.  $\pi_{g\ell}^{\overline{S_\lambda} X}$ , which must sum up to one, is the conditional probability of being in the level  $g$  of  $S_\lambda$  given the individual is in the level  $\ell$  of the latent variable (McCuthcheon, 1987; Hagenaaars, 1993, Heinen, 1996). Latent class probabilities, which must sum to one, describe the distribution of the latent variable within the observed measures. Latent class number determines the number of latent characteristics. Conditional probabilities reflect the degree of having the latent characteristic for a subject, who is in a given latent class.

#### 4. Models with Error of Measurement

These models are variants of Guttman scaling. Items are ordered from the least difficult to most difficult, thus there is one correct ordering. In Guttman scales, if once an individual responds negatively to an easier item, she will respond negatively to the other more difficult items. Therefore, Guttman scales are deterministic. The deterministic nature of Guttman scales can be improved by allowing the measurement error. These models are restricted latent class models. There are  $|X| = t + 1$  response patterns or latent classes corresponding to  $t$  binary item. In this case, some observations will respond according to  $|X| = t + 1$  response patterns and remaining will respond according to  $2^t - (t + 1)$  response patterns (McCuthcheon, 1987; Heinen, 1996). Because the items are ordered, corresponding response patterns construct an ordinal latent variable. There are four main models with error of measurement.

**Proctor's model:** Proctor's model is the first probabilistic variation of Guttman model and proposed by Proctor (1970). In the model, each scale item has error rates that are assumed to be same over all items and scale types. So, there are equality restrictions on the conditional probabilities of the scale items for each latent class.

**Item-specific error rate model:** This model relaxes the assumption that all scale items have the same error rate (Clogg and Sawyer, 1981). Instead, the

assumption that there are different error rates for each of the  $k$  items. In this models, there are equality restrictions on conditional probabilities associated with each item and latent class over the response patterns.

**True-type-specific error rate model:** True-type-specific error rate model relaxes the assumption of the Proctor's model that all scale types have the same error rate (Clogg and Sawyer, 1981). However, these models assume that incorrect response probabilities of items are the same in each latent class. To construct the model, equality restrictions are put on the conditional probabilities of scale items for each latent class.

**Lazarsfeld's latent distance model:** Latent distance model is proposed by Lazarsfeld (1950a, 1950b). The main assumption of the model is that error rates are specific to the items rather than scale types. There is the assumption that error rates for incorrect and correct responses for an item are different from each other; and this assumption is valid for all scale types except the least and most difficult ones. To set the model, the same equality constraints with the item-specific error rate model are imposed on the conditional probabilities of the first and last scale types, and for the rest, there are equality restrictions for each level of the manifest variables and corresponding latent classes on the conditional probabilities.

## 5. Log-linear Approach to the Latent Class Models

Log-linear approach linearizes the latent class model. In the log-linear representation of (2.1), the log-linear model for the incomplete table is as follows for  $N = 3$  and  $\ell = 1, \dots, |X|$ :

$$\log n_{ijkl} = \beta + \beta_i^1 + \beta_j^2 + \beta_k^3 + \beta_\ell^X + \beta_{i\ell}^{1X} + \beta_{j\ell}^{2X} + \beta_{k\ell}^{3X}. \quad (5.1)$$

Definitions of log-linear parameters are straightforward as in Section 2. For the nominal latent variable, the relation between conditional probabilities and log-linear parameters is explained over the eq. (5.1) as

$$\pi_{g\ell}^{\overline{S_\lambda X}} = \frac{\exp(\beta_g^\lambda + \beta_g^{\lambda X})}{\sum_{g=1}^{|\overline{S_\lambda}|} \exp(\beta_g^\lambda + \beta_g^{\lambda X})}, \quad (5.2)$$

where the definition of the  $g$  is the same as in eq. (3.1). Hence, conditional probabilities are represented in the form of log-linear parameters (Haberman, 1979; Heinen, 1996). By this way, restrictions on the conditional probabilities

are imposed on log-linear parameters or design matrix, and estimation process is easier than the maximum likelihood (ML) estimation of the restricted latent class analysis.

In the log-linear representation of (2.2), the log-linear model for the incomplete table is as follows for  $N = 3$  and  $\ell = 1, \dots, |X|$ :

$$\log n_{ijkl} = \beta + \beta_i^1 + \beta_j^2 + \beta_k^3 + \beta_\ell^X + \tau_{i\ell}^{1X} \nu_\ell + \tau_{j\ell}^{2X} \nu_\ell + \tau_{k\ell}^{3X} \nu_\ell. \quad (5.3)$$

Definitions of elements of log-linear model are straightforward as in Section 2.

Heinen (1996) notes that when latent variable is ordinal and manifest variables are nominal, there had not any estimation method proposed for latent class models, until 1996. In addition, we have not been coincided with any citation on the subject in the literature. In this section, a log-linear approach is introduced for this case. The log-linear model for contingency tables including an ordinal variable is given in eq. (2.2). It is the same for the ordinal latent variable case but row (column) effect models are used instead of the log-linear independence models. Interaction effects between nominal and ordinal variables are perceived as row (column) effect parameters. Under these definitions, we modify the eq. (5.2) for the case that incomplete table includes and ordinal variable corresponding to the latent variable. Over the eq. (5.3), the relation between log-linear parameters and conditional probabilities is found as follows:

$$\pi_{g\ell}^{\overline{S_\lambda X}} = \frac{\exp(\beta_g^\lambda + \tau_g^{\lambda X} \nu_\ell)}{\sum_{g=1}^{|\overline{S_\lambda}|} \exp(\beta_g^\lambda + \tau_g^{\lambda X} \nu_\ell)}. \quad (5.4)$$

The restrictions imposed on conditional probabilities are handled using (5.4) in the ordinal latent variable case over a row (column) effects log-linear model.

Models including the error of measurement contain an ordinal latent variable. Therefore, use of log linear models in the analysis of these models is possible due to the eq. (5.4). An appropriate analysis of the models with error of measurement can be made by converting the assumptions given in Section 4 to restrictions on the conditional probabilities, and expressing them as the restrictions on the log-linear parameters, namely the restrictions on the elements of design matrix.

Heinen (1996) suggests using the independence model and (5.2) imposing the restrictions on the log-linear parameters; however in this case ordinal structure of the latent variable cannot be reflected. In fact, omitting the ordinal structure of the latent variable means omitting the difficulty levels of the items.

Here choice of score values is important, because score values should reflect the ordinal structure correctly. An inappropriate choice can cause the algorithm,

which is used to obtain ML estimates, not to converge. Scores can be chosen as the proportion of the observed individuals in each latent class or integers from 1 to  $|X|$  or  $\nu_k - \bar{\nu}$ , where  $\bar{\nu}$  is the average of relevant score values.

An EM algorithm is used to obtain parameter estimates. Initial values,  $\hat{n}_{ijkl}^0$ , are determined in the E-step. Then estimated cell counts are obtained using

$$\hat{y}_{ijkl} = y_{ijk} \frac{\hat{n}_{ijkl}}{\sum_{\ell=1}^{|X|} \hat{n}_{ijkl}}.$$

When summed over the levels of latent variable, estimated observed values are equal to the observed counts. In the M-step, estimated observed counts are taken as directly observed counts and  $\hat{n}_{ijkl}$  is updated. On this step, Newton-Raphson method is used to obtain updated estimates. Then E-step is revisited. This loop is continued until the algorithm converges (Hagenaars, 1993).

## 6. Analysis of women's liberation data

Considered data set is taken from Heinen (1996, p.46). Data comes from a Dutch study of sociocultural developments in the Netherlands. Felling *et al.* (1987) give detailed information about the study. There are five binary items that

1. Women's liberation sets women against men ( $S_3$ ).
2. It's better for a wife not to have a job because that always poses problems in the household, especially if there are children ( $S_2$ ).
3. The most natural situation occurs when the man is breadwinner and the women runs the household and takes care of the children ( $S_4$ ).
4. It isn't really as important for a girl to get a good education as it is for a boy ( $S_1$ ).
5. A woman is better suited to raise small children than a man ( $S_5$ ).

For this data set,  $|S_1| = \dots = |S_5| = 2$  and  $i, j, k, m, r \in \{1, 2\}$ . Ordering of items from the most to least difficult to be agree is that  $S_1, S_2, S_3, S_4, S_5$ . Response patterns are  $(1, 1, 1, 1, 1)$ ,  $(1, 1, 1, 1, 2)$ ,  $(1, 1, 1, 2, 2)$ ,  $(1, 1, 2, 2, 2)$ ,  $(1, 2, 2, 2, 2)$ ,  $(2, 2, 2, 2, 2)$ . Each pattern is associated with a latent class. Actually, the responses on the items are rated on a 5-point scale. However, Heinen

(1996) recodes them in two categories that the category 1 includes “agree entirely”, “agree” or “natural” responses, and the category 2 includes “don’t agree” and “don’t agree at all” responses.

Women’s liberation data set is also analyzed by Heinen (1996) by means of scaling models with error of measurement assuming that the latent variable is in metric scale over independence log linear model. We reanalyze the data using our approach over models with error of measurement with row effects log linear models. The EM algorithm, mentioned in Section 5, was run until the absolute mean difference between the estimated values of the parameters was less than  $10^{-13}$ .

**Proctor’s model:** The restrictions on the conditional probabilities that make each scale item has the same error rate over all items and scale types are

$$\begin{aligned}
& \pi_{11}^{\overline{S_1}X} = \pi_{11}^{\overline{S_2}X} = \pi_{11}^{\overline{S_3}X} = \pi_{11}^{\overline{S_4}X} = \pi_{11}^{\overline{S_5}X} \\
= & \pi_{12}^{\overline{S_1}X} = \pi_{12}^{\overline{S_2}X} = \pi_{12}^{\overline{S_3}X} = \pi_{12}^{\overline{S_4}X} = \pi_{12}^{\overline{S_5}X} \\
= & \pi_{13}^{\overline{S_1}X} = \pi_{13}^{\overline{S_2}X} = \pi_{13}^{\overline{S_3}X} = \pi_{13}^{\overline{S_4}X} = \pi_{13}^{\overline{S_5}X} \\
= & \pi_{14}^{\overline{S_1}X} = \pi_{14}^{\overline{S_2}X} = \pi_{14}^{\overline{S_3}X} = \pi_{14}^{\overline{S_4}X} = \pi_{14}^{\overline{S_5}X} \\
= & \pi_{15}^{\overline{S_1}X} = \pi_{15}^{\overline{S_2}X} = \pi_{15}^{\overline{S_3}X} = \pi_{15}^{\overline{S_4}X} = \pi_{15}^{\overline{S_5}X} \\
= & \pi_{26}^{\overline{S_1}X} = \pi_{26}^{\overline{S_2}X} = \pi_{26}^{\overline{S_3}X} = \pi_{26}^{\overline{S_4}X} = \pi_{26}^{\overline{S_5}X}.
\end{aligned} \tag{6.1}$$

Heinen (1996) notes that  $G^2 = 108.62$  and  $P = 0.00$ , thus the model is not statistically significant. Here  $G^2$  is the likelihood ratio statistic (Agresti, 2002, p. 24). Each  $\pi_{\cdot}^{\overline{S}X}$  is written in log-linear parameters using (5.4) and restrictions given by (6.1) are applied to get the restrictions of Proctor’s model in log-linear parameters instead of conditional probabilities. The same way is followed for other models of interest. Therefore, it is obtained that

$$\begin{aligned}
& \beta_2^5 + \tau_2^{5X} \nu_2 = \beta_2^4 + \tau_2^{4X} \nu_3 = \beta_2^5 + \tau_2^{5X} \nu_3 \\
= & \beta_2^3 + \tau_2^{3X} \nu_4 = \beta_2^4 + \tau_2^{4X} \nu_4 = \beta_2^5 + \tau_2^{5X} \nu_4 \\
= & \beta_2^2 + \tau_2^{2X} \nu_5 = \beta_2^3 + \tau_2^{3X} \nu_5 = \beta_2^4 + \tau_2^{4X} \nu_5 \\
= & \beta_2^1 + \tau_2^{1X} \nu_6 = \beta_2^2 + \tau_2^{2X} \nu_6 = \beta_2^3 + \tau_2^{3X} \nu_6 \\
= & \beta_2^5 + \tau_2^{5X} \nu_5 = \beta_2^4 + \tau_2^{4X} \nu_6 = \beta_2^5 + \tau_2^{5X} \nu_6.
\end{aligned}$$

Here,  $|X| = 6$ , and if  $\beta_2^1 = \beta_2^2 = \beta_2^3 = \beta_2^4 = \beta_2^5$  then  $\nu_2 = \nu_3 = \nu_4 = \nu_5 = \nu_6 = c$ , where  $c$  is a constant score value. Hence, the design matrix consists one column for main effect of good education, five columns for latent classes and one column for the row effect parameter. When  $\nu_1 = 0.055$  and  $c = 0.04762$ , error rate is obtained as 0.23, which is reported as 0.13 by Heinen (1996). Obtained latent proportions are 0.054, 0.155, 0.242, 0.118, 0.145 and 0.286, which are so close to Heinen’s (1996) results. In addition, our  $G^2$  and corresponding  $P$  value are very



close to Heinen's results.

### Item-specific error rate model:

The restrictions on the conditional probabilities that assign different error rates for each of the  $k$  items are

$$\begin{array}{cccccc}
 \pi_{11}^{\overline{S_1 X}} & = \pi_{12}^{\overline{S_1 X}} & = \pi_{13}^{\overline{S_1 X}} & = \pi_{14}^{\overline{S_1 X}} & = \pi_{15}^{\overline{S_1 X}} & = \pi_{26}^{\overline{S_1 X}} \\
 \pi_{11}^{\overline{S_2 X}} & = \pi_{12}^{\overline{S_2 X}} & = \pi_{13}^{\overline{S_2 X}} & = \pi_{14}^{\overline{S_2 X}} & = \pi_{25}^{\overline{S_2 X}} & = \pi_{26}^{\overline{S_2 X}} \\
 \pi_{11}^{\overline{S_3 X}} & = \pi_{12}^{\overline{S_3 X}} & = \pi_{13}^{\overline{S_3 X}} & = \pi_{24}^{\overline{S_3 X}} & = \pi_{25}^{\overline{S_3 X}} & = \pi_{26}^{\overline{S_3 X}} \\
 \pi_{11}^{\overline{S_4 X}} & = \pi_{12}^{\overline{S_4 X}} & = \pi_{23}^{\overline{S_4 X}} & = \pi_{24}^{\overline{S_4 X}} & = \pi_{25}^{\overline{S_4 X}} & = \pi_{26}^{\overline{S_4 X}} \\
 \pi_{21}^{\overline{S_5 X}} & = \pi_{22}^{\overline{S_5 X}} & = \pi_{23}^{\overline{S_5 X}} & = \pi_{24}^{\overline{S_5 X}} & = \pi_{25}^{\overline{S_5 X}} & = \pi_{26}^{\overline{S_5 X}}.
 \end{array} \tag{6.2}$$

For this model,  $G^2 = 27.37$  and  $P = 0.159$ , when the latent variable is treated as metric (Heinen, 1996). Restrictions, given by (6.2), are expressed in terms of log-linear parameters using (5.4) as follows:

$$\begin{array}{l}
 \beta_2^1 + \tau_2^{1X} \nu_6 \\
 \beta_2^2 + \tau_2^{2X} \nu_5 = \beta_2^2 + \tau_2^{2X} \nu_6 \\
 \beta_2^3 + \tau_2^{3X} \nu_4 = \beta_2^3 + \tau_2^{3X} \nu_5 = \beta_2^3 + \tau_2^{3X} \nu_6 \\
 \beta_2^4 + \tau_2^{4X} \nu_3 = \beta_2^4 + \tau_2^{4X} \nu_4 = \beta_2^4 + \tau_2^{4X} \nu_5 = \beta_2^4 + \tau_2^{4X} \nu_6 \\
 \beta_2^5 + \tau_2^{5X} \nu_2 = \beta_2^5 + \tau_2^{5X} \nu_3 = \beta_2^5 + \tau_2^{5X} \nu_4 = \beta_2^5 + \tau_2^{5X} \nu_5 = \beta_2^5 + \tau_2^{5X} \nu_6.
 \end{array}$$

In this case, there are five columns for main effects of manifest variables, five columns for latent classes and a column for each row effect parameter in the design matrix. It is taken as  $\nu_5 = \nu_6 = c_1$ ,  $\nu_4 = \nu_5 = \nu_6 = c_2$ ,  $\nu_3 = \nu_4 = \nu_5 = \nu_6 = c_3$  and  $\nu_2 = \nu_3 = \nu_4 = \nu_5 = \nu_6 = c_4$  in the columns corresponding to  $\tau_j^{2X}$ ,  $\tau_k^{3X}$ ,  $\tau_m^{4X}$  and  $\tau_r^{5X}$ , respectively. Constants  $c_1$ ,  $c_2$ ,  $c_3$  and  $c_4$ , and implementation conditions are determined by the same manner as in Proctor's model. For this model  $G^2 = 21.88$  with  $P = 0.147$ , which are close to those noted by Heinen (1996, p.78). Five error rates are obtained as 0.0246, 0.1427, 0.4858, 0.4357, 0.5766; and latent proportions are found as 0.355, 0.091, 0.181, 0.044, 0.099, 0.230. These error rates and latent proportions are very different from those reported by Heinen (1996, p.81).

### True-type-specific error rate model

The restrictions on the conditional probabilities that make incorrect response

probabilities of items same in each latent class are

$$\begin{aligned}
\pi_{11}^{\overline{S_1 X}} &= \pi_{11}^{\overline{S_2 X}} = \pi_{11}^{\overline{S_3 X}} = \pi_{11}^{\overline{S_4 X}} = \pi_{11}^{\overline{S_5 X}} \\
\pi_{12}^{\overline{S_1 X}} &= \pi_{12}^{\overline{S_2 X}} = \pi_{12}^{\overline{S_3 X}} = \pi_{12}^{\overline{S_4 X}} = \pi_{22}^{\overline{S_5 X}} \\
\pi_{13}^{\overline{S_1 X}} &= \pi_{13}^{\overline{S_2 X}} = \pi_{13}^{\overline{S_3 X}} = \pi_{23}^{\overline{S_4 X}} = \pi_{23}^{\overline{S_5 X}} \\
\pi_{14}^{\overline{S_1 X}} &= \pi_{14}^{\overline{S_2 X}} = \pi_{24}^{\overline{S_3 X}} = \pi_{24}^{\overline{S_4 X}} = \pi_{24}^{\overline{S_5 X}} \\
\pi_{15}^{\overline{S_1 X}} &= \pi_{25}^{\overline{S_2 X}} = \pi_{25}^{\overline{S_3 X}} = \pi_{25}^{\overline{S_4 X}} = \pi_{25}^{\overline{S_6 X}} \\
\pi_{25}^{\overline{S_1 X}} &= \pi_{25}^{\overline{S_2 X}} = \pi_{25}^{\overline{S_3 X}} = \pi_{25}^{\overline{S_4 X}} = \pi_{25}^{\overline{S_6 X}}.
\end{aligned} \tag{6.3}$$

For true-type-specific error rate model, it is reported by Heinen (1996) that  $G^2 = 92.70$  and  $P = 0.00$ . When restrictions, given by (6.3), are expressed in terms of log-linear parameters using (5.4), it is obtained that

$$\begin{aligned}
\beta_2^5 + \tau_2^{5X} \nu_2 \\
\tau_2^{4X} &= \tau_2^{5X} \\
\tau_2^{3X} &= \tau_2^{4X} = \tau_2^{5X} \\
\tau_2^{2X} &= \tau_2^{3X} = \tau_2^{4X} = \tau_2^{5X} \\
\tau_2^{1X} &= \tau_2^{2X} = \tau_2^{3X} = \tau_2^{4X} = \tau_2^{5X}.
\end{aligned}$$

For this model, the design matrix of log-linear model consists one column for the main effect of  $S_5$ , five columns for latent classes and one column for the row effect parameter.  $G^2 = 397.1248$  with  $P = 0.00$  and error rates are found as 0.126, 0.217, 0.357, 0.646, 0.408 and 0.448, respectively. Latent class probabilities are obtained as 0.077, 0.162, 0.227, 0.068, 0.179 and 0.287. Our  $G^2$  is much greater than that of reported by Heinen (1996, p.78). This implies that we reject the significance of the true-type-specific error rate model more confidentially by our approach.

### Lazarsfeld's latent distance model

The restrictions on the conditional probabilities that make error rates for incorrect and correct responses for an item different from each other are

$$\begin{aligned}
\pi_{11}^{\overline{S_1 X}} &= \pi_{12}^{\overline{S_1 X}} = \pi_{13}^{\overline{S_1 X}} = \pi_{14}^{\overline{S_1 X}} = \pi_{15}^{\overline{S_1 X}} = \pi_{26}^{\overline{S_1 X}} \\
\pi_{11}^{\overline{S_2 X}} &= \pi_{12}^{\overline{S_2 X}} = \pi_{13}^{\overline{S_2 X}} = \pi_{14}^{\overline{S_2 X}} \\
\pi_{25}^{\overline{S_2 X}} &= \pi_{26}^{\overline{S_2 X}} \\
\pi_{11}^{\overline{S_3 X}} &= \pi_{12}^{\overline{S_3 X}} = \pi_{13}^{\overline{S_3 X}} \\
\pi_{24}^{\overline{S_3 X}} &= \pi_{25}^{\overline{S_3 X}} = \pi_{26}^{\overline{S_3 X}} \\
\pi_{11}^{\overline{S_4 X}} &= \pi_{12}^{\overline{S_4 X}} \\
\pi_{23}^{\overline{S_4 X}} &= \pi_{24}^{\overline{S_4 X}} = \pi_{25}^{\overline{S_4 X}} = \pi_{26}^{\overline{S_4 X}} \\
\pi_{21}^{\overline{S_5 X}} &= \pi_{22}^{\overline{S_5 X}} = \pi_{23}^{\overline{S_5 X}} = \pi_{24}^{\overline{S_5 X}} = \pi_{25}^{\overline{S_5 X}} = \pi_{26}^{\overline{S_5 X}}.
\end{aligned} \tag{6.4}$$

For the latent distance model, Heinen (1996) reports that  $G^2 = 24.02$  and  $P = 0.154$ . When restrictions, given by (6.4), are expressed in terms of log-linear parameters using (5.4), it is obtained that

$$\beta_2^1 + \tau_2^{1X} \nu_6; \nu_5 = \nu_6; \nu_4 = \nu_5 = \nu_6; \nu_3 = \nu_4 = \nu_5 = \nu_6; \nu_2 = \nu_3 = \nu_4 = \nu_5 = \nu_6.$$

The design matrix contains five columns for main effects of manifest variables and latent classes and a column for each restriction. Significance of the model is concluded by  $G^2 = 20.11$  with  $P = 0.0925$ . Error rates are 0.416, 0.027, 0.033, 0.093, 0.158, 0.008, 0.154 and 0.0193. These error rates are very different from those reported by Heinen (1996, p.80). Latent class probabilities are 0.278, 0.106, 0.156, 0.044, 0.118 and 0.297. The third and last latent class probabilities are close to those noted by Heinen (1996, p.80). The type of used log linear model is effective on the results and inferences.

In conclusion, fits of Proctor's and True-type-specific error rate models to the women's liberation data are poor, while those of Item-specific error rate and Lazarsfeld's latent distance model are statistically significant. Moreover, latent class probabilities obtained by both models are very close. The first response pattern has the greatest latent class probability according to both models. However, the highest error rates are seen for the last and first response patterns for item specific error rate and Lazarsfeld's latent distance models, respectively.

Proctor's model seems not to be effected by the type of log linear model. A possible cause of this is small number of parameters in both kinds of log linear models due to the equality restrictions of the model. For all models,  $G^2$  values and corresponding  $p$ -values are close in our and Heinen's analyses both. However, error rates and latent class probabilities are very different for all models except Proctor's model in both analyses. This shows the effect of inclusion of ordinality in the analysis. The differences are due to the loss of information when independence log linear model with metric latent variable is used.

## References

- Agresti, A. (2002). *Categorical Data Analysis*, Wiley.
- Clogg, C.C. and Sawyer, D.O. (1981). A comparison of alternative models for analyzing the scalability of response patterns. *Sociological Methodology*, Josey-Bass.
- Felling, A., Peters, J., Schreuder, O. (1987). *Religion in Dutch society 85: Documentation of a national survey on religious and secular attitudes in 1985*. Steinmetz Archive.
- Formann, A.K. (1992). Linear logistic latent class analysis for polychotomous data. *Journal of American Statistical Association* **87**, 476-486.

- Guttman, L.A. (1950). The basis for scalogram analysis. In *Measurement and Prediction* (Edited by S.A. Stouffer *et al.*), Princeton University Press.
- Haberman, S.J. (1979). *Analysis of Qualitative Data, vol 2. New Developments*. Academic Press.
- Hagenaars, J.A. (1993). *Loglinear Models With Latent Variables*. Sage.
- Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Sage.
- Lazarsfeld, P.F. (1950a). The logical and mathematical foundations of latent structure analysis. In *Measurement and Prediction* (Edited by S.A. Stouffer *et al.*), Princeton University Press.
- Lazarsfeld, P.F. (1950b). The interpretation and computation of some latent structures. In *Measurement and Prediction* (Edited by S.A. Stouffer *et al.*), Princeton University Press.
- Lazarsfeld, P.F., Henry, N.W. (1968). *Latent Structure Analysis*. Houghton Mifflin.
- McCutcheon, A.L. (1987). *Latent Class Analysis*. Sage.
- Proctor, C.H. (1970). A probabilistic formulation and statistical analysis of Guttman scaling. *Psychometrika* **35**, 73-78.

Received March 19, 2009; accepted May 6, 2009.

Haydar Demirhan  
Department of Statistics  
Hacettepe University  
Beytepe, Ankara, 06800, TURKEY  
haydarde@hacettepe.edu.tr