國 立 交 通 大 學

統計學研究所

碩 士 論 文

藉由K均值分群與分裂式階層分群程序預測潛在群體

Prediction of Underlying Latent Classes via K-means and Divisive Hierarchical Procedures

研 究 生：許仲竹

指導教授：黃冠華 博士

中 華 民 國 九 十 七 年 六 月

藉由K均值分群與分裂式階層分群程序預測潛在群體

Prediction of Underlying Latent Classes via K-means and Divisive Hierarchical Procedures

研　究　生：許仲竹　　　Student: Chung-Chu Hsu

指導教授：黃冠華　　　Advisor: Dr. Guan-Hua Huang

國　立　交　通　大　學

統計學研究所

碩　士　論　文

A Thesis
Submitted to institute of Statistics
College of Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Master
in
Statistics
June 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年六月

藉由K均值分群與分裂式階層分群程序預測潛在群體

研究生：許仲竹　　　　指導教授：黃冠華　博士

國立交通大學統計學研究所

摘要

本研究的主要目的是藉由群聚分析的方法預測潛在群體。利用群聚方法中的 k 均值分群和分裂階層分群的想法，將原本的距離測度改為相關係數或共變異數，對所有的主體分群，使得屬於同一群的主體所測得的各項目能互相獨立。利用模擬來評估參數估計的表現，除此之外，還利用精神分裂症和乳癌的微陣列資料為例，作更詳細的說明。模擬結果顯示，k 均值分群法所估出來的參數都相當靠近真實的參數，但是分裂式階層分群法表現得並不好;然而，在乳癌資料的例子裡，分裂階層分群法成功的將主體分群，也對潛在群體做了不錯的預測。

關鍵字：潛在群體、K 均值分群、分裂式分群

Prediction of Underlying Latent Classes via K-means and Divisive Hierarchical Procedures

Student: Chung-Chu Hsu   Advisor: Dr. Guan-Hua Huang

Institute of Statistics
National Chiao Tung University

ABSTRACT

The aim of the study is to predict the underlying latent class via k-means and divisive hierarchical clustering methods. We use the correlation (or covariance) among items as the distance measure to group objects such that, for all objects who belong to the same latent class, items are ”independent”. A simulation study is presented to evaluate the behavior of estimating parameters. Besides, the schizophrenia and breast cancer microarray data were used for illustration. The results of the simulation studies displayed that the estimated parameters by k-means method are closed to the true parameters, but the divisive hierarchical method didn’t perform well. However, the divisive hierarchical approach makes the successful division and predicts the latent class membership well for breast cancer data.

Key words: Regression extend of latent class analysis, k-means, divisive hierarchical

誌謝

最重要的要感謝我的指導教授黃冠華老師，這一年在老師的教導下，才能使我完成這篇論文，不僅如此，也引導我學習如何解決問題，使我在辛苦的研究過程中找到樂趣。

當然我同學及朋友的鼓勵也是我這一年做研究的動力。感謝跟我同一家的阿耕、彥銘師兄及佩芳師姐，有事沒事一起吐苦水、交換意見，互相鼓勵；打球的夥伴們，讓我在疲憊的研究所生活得到充電的機會，還有所上的每一位老師、同學，感謝你們的照顧，讓我在這兩年中，不僅得的了許多的知識，更擁有珍貴的友誼，真的很開心認識你們。

黃冠華老師、邱燕楓教授、陳君厚教授及洪志真老師，也要感謝你們寶貴的意見，讓我的論文更佳的完善。

即將要畢業了，我將帶著感謝及美好的回憶離開，進交大、離開交大，這短短的兩年，但卻有我無限的回憶，感謝你們的支持及關愛，現在的我才能有無比的信心及勇氣迎接下一個階段的挑戰。

最後，僅將此論文送給我愛的父母及老師及我身邊的同學們，我想這是我們大家努力的結果，也將一同分享論文完成的喜悅，感謝大家。

許仲竹 謹 誌于

國立交通大學統計研究所

中華民國九十七年七月三日

Contents

Abstract (in Chinese)

Abstract (in Chinese) i

Abstract (in English) ii

Acknowledgements (in Chinese) iii

Contents iv

List of Tables v

List of Figures viii

List of Tables....v

1. Introduction....1

2. Literature review....3

2.1 Latent class analysis (LCA) 3

2.2 Regression extension of latent class analysis (RLCA) 3

2.3 Marginalization of the regression extension of latent class model 5

2.3.1 Marginalizing the covariate effects on conditional probabilities 6

2.3.2 Marginalizing the covariate effects on latent prevalences 7

2.4 Hierarchical clustering methods....8

2.5 Ward's hierarchical clustering method....9

2.6 *K*-means method....10

3. Models....12

3.1 LCA 12

3.2 RLCA 12

4. Parameter estimations by clustering analysis....15

4.1 Latent class membership estimations for LCA 15

4.2 Latent class membership estimations for RLCA 21

4.3 Parameter estimation by viewing estimated latent class as known variable....22

5. Simulation study....24

5.1 Generated data from RLCA model....24

5.2 Simulation results....25

6. Example....30

7. Discussion....39

References....40

List of Tables

Table 1: Values of α_0 and α_{Lm} in 3 class case..42

Table 2: Values of β_0 and β_{pj} in 3 class case..42

Table 3: Values of α_0 and α_{Lm} in 6 class case..43

Table 4: Values of β_0 and β_{pj} in 6 class case..43

Table 5: Values of α_0 and α_{Lm} in 2 class case..44

Table 6: Values of β_0 and β_{pj} in 2 class case..44

Table 7: Average parameters estimations for 100 replication in 3-class model,N=100 45

Table 8: Average parameters estimations for 100 replication in 3-class model, N=100..46

Table 9: Average conditional Probability for 100 replication in 3-class model, N=100..46

Table 10: Average Latent Prevalence for 100 replication in 3-class model, N=100.48

Table 11: Average Correlation Coefficients for 100 replication in3-class model,N=100..48

Table 12: Average Match Proportions for 100 replication in 3-class model, N=100 48

Table 13: Average parameters estimations for 100 replication in 3-class model, N=500..49

Table 14: Average parameters estimations for 100 replication in 3-class model, N=500..50

Table 15: Average conditional Probability for 100 replication in 3-class model, N=500..50

Table 16: Average Latent Prevalences for 100 replication in 3-class model, N=500 52

Table 17: Average Correlation Coefficients for 100 replication in 3-class model, N=500..52

Table 18: Average Match Proportions for 100 replication in 3-class model, N=500 52

Table 19: Average parameters estimations for 100 replication in 6-class model, N=300..53

Table 20: Average parameters estimations for 100 replication in 6-class model, N=300..55

Table 21: Average conditional Probability for 100 replication in 6-class model, N=300..56
Table 22: Average Latent Prevalences for 100 replication in 6-class model, N=300 60
Table 23: Average Correlation Coefficients for 100 replication in 6-class model, N=300 (total number of not NA values in parentheses) 61
Table 24: Average Match Proportions for 100 replication in 6-class model, N=300 61
Table 25: Average parameters estimations for 100 replication in 6-class model, N=1000 (standard error in multinomial regression / sample standard error for 100 replication) 62
Table 26: Average parameters estimations for 100 replication in 6-class model, N=1000..65
Table 27: Average conditional Probability for 100 replication in 6-class model, N=1000..66
Table 28: Average Latent Prevalences for 100 replication in 6-class model, N=1000 70
Table 29: Average Correlation Coefficients for 100 replication in 6-class model, N=1000 (total number of not NA values in parentheses) 70
Table 30: Average Match Proportions for 100 replication in 6-class model, N=1000 70
Table 31: Average parameters estimations for 100 replication in 2-class model, N=150..71
Table 32: Average parameters estimations for 100 replication in 2-class model, N=150..72
Table 33: Average conditional Probability for 100 replication in 2-class model, N=150..73
Table 34: Average Latent Prevalences for 100 replication in 2-class model, N=150 74
Table 35: Average Correlation Coefficients for 100 replication in 2-class model, N=150 (total number of not NA values in parentheses) 74
Table 36: Average Match Proportions for 100 replication in 2-class model, N=150 74
Table 37: Average parameters estimations for 100 replication in 2-class model, N=700..75
Table 38: Average parameters estimations for 100 replication in 2-class model, N=700..76
Table 39: Average conditional Probability for 100 replication in 2-class model,

N=700 77
Table 40: Average Latent Prevalences for 100 replication in 2-class model, N=700 78
Table 41: Average Correlation Coefficients for 100 replication in 2-class model, N=700 (total number of not NA values in parentheses) 78
Table 42: Average Match Proportions for 100 replication in 2 class model, N=700 78
Table 43: Composition of classes of patients at the acute state by the four-class RLCA model with divisive hierarchical clustering method 79
Table 44: External validity of classes of patients at the acute state by the four-class RLCA model with divisive hierarchical clustering method 80
Table 45: Composition of classes of patients at the subsided state by the three-class RLCA model with divisive hierarchical clustering method 81
Table 46: External validity of classes of patients at the subsided state by the three-class RLCA model with divisive hierarchical clustering method 82
Table 47: Composition of classes of patients at the acute state by the four-class RLCA model with k-means clustering method 83
Table 48: External validity of classes of patients at the acute state by the four-class RLCA model with k-means clustering method 84
Table 49: Composition of classes of patients at the subsided state by the three-class RLCA model with k-means clustering method 85
Table 50: External validity of classes of patients at the subsided state by the three-class RLCA model with k-means clustering method 86
Table 51: External validity of classes of breast cancer patients by the two-class RLCA model with divisive hierarchical clustering method 87
Table 52: Predictions of class membership of 19 tumours by divisive hierarchical clustering method 87
Table 53: External validity of classes of breast cancer patients by the two-class RLCA model with k-means clustering method 88
Table 54: Predictions of class membership of 19 tumours by k-means clustering method 88

List of Figures

Figure 1: An example of k-means algorithm procedure. 18
Figure 2: An example of divisive hierarchical algorithm procedure. 20
Figure 3: Heatmap for schizophrenia patients at the acute state 89
Figure 4: Heatmap for schizophrenia patients at the subsided state 90
Figure 5: Heatmap for breast tumours 91

1. Introduction

Latent class analysis (LCA), originally described by Green (1951) and systematically developed by Lazarsfeld and Henry (1968), Goodman (1974), has been found useful for classifying objects based on their responses to a set of categorical items. Latent class models have proven useful for analyzing relationships between measured multiple indicators and covariates of interest. Such models summarize shared features of the multiple indicators as an underlying categorical variable, and the indicators' substantive associations with predictors are built directly and indirectly in unique model parameters.

The basic model postulates an underlying categorical latent variable, say, J categories, and measured items are assumed independent of one another within any category of the latent variable. Observed relationships among measured variables are thus assumed to result from the underlying classification of the data produced by the categorical latent variable.

Latent class analysis may legitimately be viewed as the analog of cluster analysis. The term *cluster analysis* encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. In this research, instead of grouping objects of "similar kind" into respective categories, we apply the divisive hierarchical ideas of clustering methods with the correlation among items as the distance measure to group objects such that, for all objects who belong to the same latent class, items are "independent".

Recently several authors extended the LCA model to describe the effects of measured covariates on the underlying mechanism (Dayton and Macready, 1988; Vander Heijden, Dessens and Bökenholt, 1996; Bandeen-Roche, Migliorette, Zeger and Rathouz, 1997), or on measured item distributions within latent levels (Melton, Liang and Pulver, 1994). These extended LCA models are called the regression extension of latent class analysis (RLCA) models. For the RLCA model, by using the marginalizing techniques to eliminate covariate effects from both the latent variable and measured indicators (Huang, 2005), our clustering idea can be also applied to the reduced LCA model to estimate the latent class membership. By viewing the latent variable as known predictors, it becomes easy to estimate the parameters in the RLCA model.

2. Literature review

2.1 Latent class analysis (LCA)

The starting point for the methodology that we let $\mathbf{Y}_i = (Y_{i1},...,Y_{iM})^T$ denote a set of M observable polytomous indicators for the ith individual in a study sample of N persons Y_{im}, $m = 1,...,M$ can take values $\{1,...,K_m\}$, where $K_m \geq 2$. The basic model postulates an underlying categorical latent variable $S_i = 1,...,J$ for individual i; within any category of the latent variable, the measured indicators are assumed to be independent of one another. Therefore, the distribution for $\mathbf{Y}_i$ can be expressed as

$$\Pr(Y_{i1} = y_1,...,Y_{iM} = y_m) = \sum_{j=1}^{J} \left\{ \eta_j \prod_{m=1}^{M} \prod_{k=1}^{K_m} p_{mkj}^{y_{mk}} \right\}, \tag{2.1}$$

where $y_{mk} = I(y_m = k) = 1$ if $y_{mk} = k$; 0 otherwise. The LCA model assumes that

$$\eta_j = \Pr(S_i = j) \text{ and } p_{mkj} = \Pr(Y_{im} = k \mid S_i = j), \tag{2.2}$$

$$i = 1,...,N \,; m = 1,...,M \,; k = 1,...,K_m \,; j = 1,...,J \,.$$

The model treats class membership probabilities, η_j, and item response probabilities conditional on class membership, p_{mkj}, as homogeneous over individuals. Heuristically, η_j is the population prevalence of class j, and p_{mkj} is the probability of an individual in class j being at levels k of Y_{im}. Goodman (1974) provided an excellent overview of the LCA model, including a maximum likelihood strategy for estimating model parameters, conditions to determine local model identifiability, a strategy to test overall model fit, and the use of constraints to identify models.

2.2 Regression extension of latent class analysis (RLCA)

Huang and Bandeen-Roche (2004) extend the latent class analysis to allow both the probabilities of latent class membership and the distribution of observed responses given latent class membership to be functionally related to concomitant variables, while preserving model identifiability. By allowing covariate effects on latent class probabilities, we summarize the effect of risk factors on the underlying mechanism. In the case of incorporation covariates into conditional probabilities, we can adjust for characteristics that determine responses other than underlying classes, hence hopefully improving the accuracy of classifying individuals. For instance, in evaluating functional disability, some data have suggested that women tend to rate tasks as “difficult” more readily than men independently of ability (Bandeen-Roche, Huang, Munoz, & Rubin, 1999). Without adjusting for the gender effect, the model might well classify some men and women with identical underlying functioning differently (men as “able”, women as “disabled”).

Let $(\mathbf{x}_{\mathrm{i}}, \mathbf{z}_i)$ be the concomitant covariates of the ith person, where $\mathbf{x}_{\mathrm{i}} = (1, x_{i1}, \ldots, x_{ip})^T$ are primary covariate hypothesized to be associated with latent class membership, S_i, and $\mathbf{z}_{\mathrm{i}} = (z_{i1}, \ldots, z_{iM})^T$ with $\mathbf{z}_{\mathrm{im}} = (1, z_{im1}, \ldots, x_{imL})^T$, $m = 1, \ldots, M$, are secondary covariates used to build direct effects on measured indicators. The sets of covariates may include any combination of continuous and discrete measures, and two sets of covariates may be mutually exclusive or overlap. The regression extension of LCA may then be stated as follows:

$$\Pr(Y_{i1} = y_1, \ldots, Y_{iM} = y_m \mid \mathbf{x}_{\mathrm{i}}, \mathbf{z}_i) = \sum_{j=1}^{J} \left\{ \eta_j(\mathbf{x}_i^T \boldsymbol{\beta}) \prod_{m=1}^{M} \prod_{k=1}^{K_m} p_{mkj}^{y_{mk}} (\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T \boldsymbol{\alpha}_m) \right\} \tag{2.3}$$

with $\eta_j(\mathbf{x}_i^T \boldsymbol{\beta})$ and $p_{mkj}^{y_{mk}} (\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T \boldsymbol{\alpha}_m)$ defined as in the generalized linear framework (McCullagh and Nelder, 1989). Often, (3) is implemented assuming generalized logit (Agresti, 1984) link functions:

$$\log\left[\frac{\eta_j(\mathbf{x}_i^T \boldsymbol{\beta})}{\eta_J(\mathbf{x}_i^T \boldsymbol{\beta})}\right] = \beta_{0j} + \beta_{1j} x_{i1} + \cdots + \beta_{pj} x_{ip} \quad \text{for } i = 1, \ldots, N;\ j = 1, \ldots, J-1 \tag{2.4}$$

and

$$\log\left[\frac{p_{mkj}\left(\boldsymbol{\gamma}_{mj}+\mathbf{z}_{im}^{T}\boldsymbol{\alpha}_{m}\right)}{p_{mKj}\left(\boldsymbol{\gamma}_{mj}+\mathbf{z}_{im}^{T}\boldsymbol{\alpha}_{m}\right)}\right]=\gamma_{mkj}+\alpha_{1mk}z_{im1}+\cdots+\alpha_{Lmk}z_{imL}$$

$$\text{for } i=1,\ldots,N;\ m=1,\ldots,M;\ k=1,\ldots,(K_m-1);\ j=1,\ldots,J. \qquad (2.5)$$

Notice that in the conditional probability model (2.5), we allow unrestricted intercepts and level-and item-specific covariate coefficients, but the coefficients vary across classes is unallowable (i.e., α_{qmk} is dependent on m, k but independent of j). This constraint is reasonable if the primary purpose of modeling conditional probabilities is to prevent possible misclassification by adjusting for characteristics associated with item measurements. It is also necessary to unambiguously distinguish covariate effects on measured response probabilities from covariate effects on class probabilities. Three assumptions complete (2.3):

(C1) $\Pr\left(Y_{i1}=y_1,\ldots,Y_{iM}=y_m \mid S_i,\mathbf{x}_i,\mathbf{z}_i\right)=\Pr\left(Y_{i1}=y_1,\ldots,Y_{iM}=y_m \mid S_i,\mathbf{z}_i\right)$;

(C2) $\Pr\left(S_i=j \mid \mathbf{x}_i,\mathbf{z}_i\right)=\Pr\left(S_i=j \mid \mathbf{x}_i\right)$;

(C3) $\Pr\left(Y_{i1}=y_1,\ldots,Y_{iM}=y_m \mid S_i,\mathbf{z}_i\right)=\prod_{m=1}^{M}\Pr\left(Y_{im}=y_m \mid S_i,\mathbf{z}_{im}\right)$.

Huang and Bandeen-Roche (2004) provided an excellent overview of the RLCA model, including model identification, Expectation-Maximization algorithm for parameter estimation, standard error calculation, convergent properties, and comparison of the RLCA model with models underlying existing latent class modeling software.

2.3 Marginalization of the regression extension of latent class model

Now we introduce a process to "eliminate" the covariates effect, hence "marginalize" the RLCA model (2.3). The marginalization process (Huang 2005) includes two stages. Stage 1 aims to eliminate $\mathbf{z}_i$ effect. And stage 2, we apply the marginalization property, proposed by Bandeen-Roche et al. (1997, to average $\mathbf{x}_i$ effect out of the latent prevalence).

2.3.1 Marginalizing the covariate effects on conditional probabilities

The key to marginalizing over $\mathbf{z}_i$ is that the process must yield random variables that follow a finite mixture distribution that is both independent of $\mathbf{z}_i$ and has J mixing components. One strategy for achieving such marginalization can be motivated by the properties of added variable plots for linear regression models.

Consider the linear model

$$\mathbf{Y} = \mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_1 + \mathbf{x}_2^{\mathrm{T}}\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \tag{2.6}$$

where $\boldsymbol{\varepsilon}$ with mean $\mathbf{0}$ and variance matrix $\mathbf{V}$. Let $\tilde{\mathbf{Y}}$ denote the residuals of regressing $\mathbf{Y}$ on $\mathbf{x}_2$, and $\mathbf{W} = \mathbf{V}^{-1}$ be the weight matrix. Then, it is well known that if $\mathbf{x}_1$ and $\mathbf{x}_2$ are orthogonal (i.e., $\mathbf{x}_1\mathbf{W}\mathbf{x}_2^{\mathrm{T}} = 0$), $\tilde{\mathbf{Y}}$ has mean $\mathbf{x}_1^{\mathrm{T}}\boldsymbol{\beta}_1$ and variance $\mathbf{V}$. Hence, the simple linear regression of $\tilde{\mathbf{Y}}$ on $\mathbf{x}_1$ yields exactly the same inferences about $\boldsymbol{\beta}_1$ as if we performed the analysis on the more complicated model (2.6) (Cook and Weisberg, 1982). Viewing the just-described stability of $\boldsymbol{\beta}_1$ as analogous to the desired stability of latent class dimension, J, the added variable property can be applied to model (2.6) to obtain the marginalized conditional probabilities.

To present the key ideas more clearly, the measured indicators $(Y_{i1},\ldots,Y_{iM})$ are assumed to be binary (i.e., $\mathrm{K}_1 = \cdots = K_M = 2$). To make the analogy to (2.6), notice that (2.5) can be viewed as fitting a logistic regression of Y_{im} on S_{i} adjusting for $\mathbf{z}_{\mathrm{im}}$, separately for each m. Let $S_{\mathrm{ij}} = \mathrm{I}(S_{\mathrm{i}} = j)$ for $i = 1,\ldots,N$; $j = 1,\ldots,J-1$. We can reparameterize (2.5) as

$$\operatorname{logit}\left[\mathrm{E}\left(Y_{im} \mid \mathbf{S}_i, \mathbf{Z}_{im}^c\right)\right] = \mathbf{S}_i^{\mathrm{T}}\boldsymbol{\gamma}_m + \left(\mathbf{Z}_{im}^c\right)^{\mathrm{T}}\boldsymbol{\alpha}_m \quad \text{for} \quad i = 1,\ldots,N;\, m = 1,\ldots,M \tag{2.7}$$

where $\mathbf{S}_{\mathrm{i}} = \left[1, S_{i1},\ldots,S_{i(J-1)}\right]^{\mathrm{T}}$;

$\mathbf{Z}_{im}^{c} = [(z_{im1} - \bar{z}_{m1}),...,(z_{imL} - \bar{z}_{mL})]^{\mathrm{T}}$, ("centered" covariate vector);

$\bar{z}_{mp} = (1/N)\sum_{i=1}^{N} z_{imp}$;

$\boldsymbol{\gamma}_{m} = [\gamma_{m0}, \gamma_{m1},..., \gamma_{m(J-1)}]^{\mathrm{T}}$; and $\boldsymbol{\alpha}_{m} = [\alpha_{1m}, \alpha_{2m},..., \alpha_{Lm}]^{\mathrm{T}}$.

Therefore, for any realization of $\mathbf{S}_{\mathrm{i}}$, (2.7) is a logistic regression with dependent variable: Y_{im} and predictors: $\mathbf{S}_{\mathrm{i}}, \mathbf{Z}_{im}^{c}$.

Next, the problem becomes how to calculate residuals form the generalized linear model

$$\mathrm{logit}\left[\mathrm{E}\left(Y_{im} \mid \mathbf{S}_{i}, \mathbf{Z}_{im}^{c}\right)\right] = \left(\mathbf{Z}_{im}^{c}\right)^{\mathrm{T}} \boldsymbol{\alpha}_{m}^{*} \quad \text{for} \quad i = 1,..., N \,;\, m = 1,..., M \tag{2.8}$$

The "pseudo-residuals" are given by

$$\mathbf{R}_{m} = [R_{1m},..., R_{Nm}]^{\mathrm{T}} = \hat{\mathbf{V}}_{m}^{-1}(\mathbf{Y}_{m} - \hat{\boldsymbol{\mu}}_{m}). \tag{2.9}$$

Here "hat" represents the estimated values;

$\mathbf{Y}_{m} = [Y_{1m},..., Y_{Nm}]^{\mathrm{T}}$; $\mathbf{V}_{m} = \mathrm{diag}(V_{1m},..., V_{Nm})$; $V_{im} = \mathrm{Var}(Y_{im})$; $\mathbf{Z}_{m}^{c} = [\mathbf{Z}_{1m}^{c},..., \mathbf{Z}_{Nm}^{c}]$

If $\mathbf{x}_{i}$ and $\mathbf{z}_{im}$ are independent, we can extract the $\mathbf{Z}_{im}^{c}$ from conditional probabilities by treating the residuals form the model (2.8) as new response variables and regressing them on $\mathbf{S}_{\mathrm{i}}$. We substitute the estimate of $\boldsymbol{\gamma}_{m}^{*}$ in the linear model

$$R_{im} = \mathbf{S}_{i}^{T} \boldsymbol{\gamma}_{m}^{*} + \varepsilon_{im}, \quad i = 1,..., N \,;\, m = 1,..., M \tag{2.10}$$

For the estimate of $\boldsymbol{\gamma}_{m}$ in the model (2.7). A formal justification shows that $\boldsymbol{\gamma}_{m}^{*}$ and $\boldsymbol{\gamma}_{m}$ can be very close under reasonable regularities. The above results can be extended to the cases where $(Y_{i1},..., Y_{iM})$ is polytomous as in (2.1) and (2.3).

2.3.2 Marginalizing the covariate effects on latent prevalences

The marginalization of model (2.3) over $\mathbf{z}_{i}$, we possesses the nice property that the covariates associated with latent class prevalences, $\mathbf{x}_i$, can be ignored.

2.4 Hierarchical clustering methods

Hierarchical clustering techniques proceed by either a series of successive mergers or a series of successive divisions. Agglomerative hierarchical methods start with the individual objects. Thus, there are initially as many clusters as objects. The most similar objects are first grouped, and these initial groups are merged according to their similarities. Eventually, as the similarity decreases, all subgroups are fused into a single cluster.

Divisive hierarchical methods work in the opposite direction. An initial single group of object is divided into two subgroups such that the objects in one subgroup are "far from" the objects in the other. These subgroups are then further divided into dissimilar subgroups; the process continues until there are as many subgroups as objects – that is, until each object forms a group.

The results of both agglomerative and divisive methods may be displayed in the form of a two-dimensional diagram known as a dendrogram. As we shall see, the dendrogram illustrates the mergers or divisions that have been made at successive levels.

In this research, we focus on divisive hierarchical procedures. We will use an algorithm based on the proposal of Macnaughton-Smith et al. (1964).Here we illustrate the divisive analysis algorithm for grouping N objects.

1. All objects as a single cluster and an $N \times N$ symmetric distance (or dissimilarities) matrix $D = \{d_{ij}\}$.
2. Looking for the object for which the dissimilarity to all other objects is largest. (If there are two such objects, we pick one at random.) This object is chosen to

initiate so-called *splinter group*.

3. For each objects of the larger group, we compute the dissimilarity with the remaining objects, and compare it to the dissimilarity with the objects of the splinter group. We choose the object which has the largest difference dissimilarity between the remaining objects with the splinter group to move into the splinter group.
4. Repeating step 3 until all the differences have become negative. Therefore, no further moves are made. The process stops and we have completed the first divisive step.
5. Then, we divide the biggest cluster, that is, the cluster with the largest diameter. (The diameter of a cluster is just the largest dissimilarity between two of its objects.) Therefore, the above procedure will be applied until all objects in a single cluster.

2.5 Ward's hierarchical clustering method

Ward (1963) considered hierarchical clustering procedures based on minimizing the "loss of information" from joining two groups. This method is usually implemented with loss of information taken to be an increase in an error sum of squares criterion, **ESS**. First, for a given cluster k, let $\mathbf{ESS}_k$ be the sum of the squared deviations of every item in the cluster from the cluster mean (centroid). If there are currently K clusters, define **ESS** as the sum of the $\mathbf{ESS}_k$ or $\mathbf{ESS} = \mathbf{ESS}_1 + \mathbf{ESS}_2 + ... + \mathbf{ESS}_\mathbf{K}$. At each step in the analysis, the union of every possible pair of clusters is considered, and the two clusters whose combination results in the smallest increase in **ESS** (minimum loss of information) are joined. Initially, each cluster consists of a single item, and, if there are N items, $\mathbf{ESS}_k = 0$, $k = 1,2,...N$, so $\mathbf{ESS} = 0$. At the other extreme, when all the clusters are combined in a single group of N items, the value of **ESS** is given by

$$\mathbf{ESS} = \sum_{j=1}^{N} \left(\mathbf{x_j} - \overline{\mathbf{x}}\right)^{\mathrm{T}} \left(\mathbf{x_j} - \overline{\mathbf{x}}\right)$$

where $\mathbf{x}_j$ is the multivariate measurement associated with the jth item and $\overline{\mathbf{x}}$ is the mean of all the items.

The results of Ward's method can be displayed as a dendrogram. The vertical axis gives the values of **ESS** at which the mergers occur.

Ward's method is based on the notion that the clusters of multivariate observations are expected to be roughly elliptically shaped. It is a hierarchical precursor to nonhierarchical clustering methods that optimize some criterion for dividing data into a given number of elliptical groups.

2.6 *K*-means method

MacQueen (1967) suggests the term *K-means* for describing an algorithm of his that assigns each item to the cluster having the nearest centroid (mean). In its simplest version, the process is composed of these three steps:

1. Partition the items into K initial clusters.

2. Proceed through the list of items, assigning an item to the cluster whose centroid (mean) is nearest. (Distance is usually computed using Euclidean distance with either standardized or unstandardized observations.) Recalculate the centroid for the cluster receiving the new item and for the cluster losing the item.

3. Repeat Step 2 until no more reassignments take place.

Rather than starting with a partition of all items into K preliminary groups in Step 1, we could specify K initial centroids (seed points) and then proceed to Step 2.

The final assignment of items to clusters will be, to some extent, dependent upon the initial partition or the initial selection of seed points. Experience suggests that most major changes in assignment occur with the first reallocation step.

3. Models

3.1 LCA

Let $(Y_{i1},...,Y_{iM})$ denote a set of M observable polytomous outcome indicators and S_i denote the unobservable class membership, for the ith individual in a study sample of N persons. Y_{im} can take values $\{1,...,K_m\}$, where $K_m \geq 2$, $m = 1,...,M$, and S_i can take values $\{1,...,J\}$. The latent class analysis model is based on the concept of conditional independence in the sense that the observed variables are assumed to be statistically independent within latent classes. Therefore, the distribution for $(Y_{i1},...,Y_{iM})$ can be expressed as the finite mixture density:

$$\Pr(Y_{i1} = y_1,...,Y_{iM} = y_m) = \sum_{j=1}^{J} \left\{ \Pr(S_i = j) \prod_{m=1}^{M} \prod_{k=1}^{K_m} [\Pr(Y_{im} = k \mid S_i = j)]^{y_{mk}} \right\}, \quad (3.1)$$

where $y_{mk} = 1$ if $y_m = k$; 0 otherwise. The LCA model assumes that $\Pr(Y_{im} = k \mid S_i = j) = p_{mkj}$, $\Pr(S_i = j) = \eta_j$, $i = 1,...,N$; $m = 1,...,M$; $k = 1,...,K_m$; $j = 1,...,J$. Thus, the model treats class membership probabilities, η_j, and item response probabilities conditional on class membership, p_{mkj}, as homogeneous over individuals. Heuristically, η_j is the population prevalence of class j, and p_{mkj} is the probability of an individual in class j being at levels k of Y_{im}.

For more detail on identifiability, parameter estimations and the test overall model fit, readers may reference Goodman (1974).

3.2 RLCA

To incorporate covariate effects into LCA, let $(\mathbf{x}_i, \mathbf{z}_i)$ be the concomitant covariates of the ith person, where $\mathbf{x}_i = (1, x_{i1},...,x_{ip})^T$ are primary covariate hypothesized to be associated with latent class membership, S_i , and $\mathbf{z}_\mathrm{i} = (z_{i1},\ldots,z_{iM})^T$ with $\mathbf{z}_\mathrm{im} = (1, z_{im1},\ldots,x_{imL})^T$, $m = 1,\ldots,M$, are secondary covariates used to build direct effects on measured indicators. The sets of covariates may include any combination of continuous and discrete measures. To marginalize the RLCA model, we begin by assuming that the two sets of covariates are mutually independent. The basic RLCA equation can be stated as

$$\Pr(Y_{i1} = y_1,\ldots,Y_{iM} = y_m \mid \mathbf{x}_\mathrm{i}, \mathbf{z}_i) = \sum_{j=1}^{J}\left\{\eta_j\left(\mathbf{x}_i^T\boldsymbol{\beta}\right)\prod_{m=1}^{M}\prod_{k=1}^{K_m} p_{mkj}^{y_{mk}}\left(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m\right)\right\} \quad (3.2)$$

with $\eta_j\left(\mathbf{x}_i^T\boldsymbol{\beta}\right)$ and $p_{mkj}^{y_{mk}}\left(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m\right)$ defined as in the generalized linear framework (McCullagh and Nelder, 1989). Often, (3.2) is implemented assuming generalized logit (Agresti, 1984) link functions:

$$\log\left[\frac{\eta_j\left(\mathbf{x}_i^T\boldsymbol{\beta}\right)}{\eta_J\left(\mathbf{x}_i^T\boldsymbol{\beta}\right)}\right] = \beta_{0j} + \beta_{1j}x_{i1} + \cdots + \beta_{pj}x_{ip} \quad \text{for} \quad i = 1,\ldots,N\,;\ j = 1,\ldots,J-1, \quad (3.3)$$

and

$$\log\left[\frac{p_{mkj}\left(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m\right)}{p_{mKj}\left(\boldsymbol{\gamma}_{mj} + \mathbf{z}_{im}^T\boldsymbol{\alpha}_m\right)}\right] = \gamma_{mkj} + \alpha_{1mk}z_{im1} + \cdots + \alpha_{Lmk}z_{imL}$$

$$\text{for} \quad i = 1,\ldots,N\,;\ m = 1,\ldots,M\,;\ k = 1,\ldots,(K_m - 1)\,;\ j = 1,\ldots,J\,. \quad (3.4)$$

If the regression coefficients in (3.3) or (3.4) are set as 0, model (3.2) reduced to models studied by Melton, Liang and Pulver (1994), Dayton and Macready (1998) or an ordinary latent class analysis (3.1).

Notice that in the conditional probability model (3.4), we allow unrestricted intercepts and level-and item-specific covariate coefficients, but we do not allow the coefficients to vary across classes (i.e., α_{qmk} is dependent on *m*, *k* but independent of j). This constraint is logical if the primary purpose of modeling conditional probabilities is to prevent possible misclassification by adjusting for characteristics associated with item measurements. It is also necessary to unambiguously distinguish covariate effects on measured response probabilities from covariate effects on class probabilities. Three assumptions complete (3.2):

(C1) $\Pr(Y_{i1} = y_1, \ldots, Y_{iM} = y_m \mid S_i, \mathbf{x}_i, \mathbf{z}_i) = \Pr(Y_{i1} = y_1, \ldots, Y_{iM} = y_m \mid S_i, \mathbf{z}_i)$;

(C2) $\Pr(S_i = j \mid \mathbf{x}_i, \mathbf{z}_i) = \Pr(S_i = j \mid \mathbf{x}_i)$;

(C3) $\Pr(Y_{i1} = y_1, \ldots, Y_{iM} = y_m \mid S_i, \mathbf{z}_i) = \prod_{m=1}^{M} \Pr(Y_{im} = y_m \mid S_i, \mathbf{z}_{im})$.

For more detail on model assumptions, identifiability and parameter estimations, readers may reference Huang and Bandeen-Roche (2004).

4. Parameter estimations by clustering analysis

The parameters in (3.2) are typically estimated by maximum likelihood (ML) for a fixed number of classes, *J*. Viewing the class membership S_i as unobservable, the LCA model (3.1) and RLCA model (3.2) becomes a typical incomplete-data problem.

Goodman (1974) provided an excellent maximum likelihood strategy for estimating model parameters in (3.1), and Huang and Bandeen-Roche (2004) had successfully used the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1997) to computing ML estimates of the parameters in (3.2) and created a powerful computer module to implement the proposed latent class model (3.2). However implementing the EM algorithm to estimate parameters in finite-mixture models is typically time-consuming. Therefore we propose an alternative clustering analysis strategy to predict parameters in (3.1) and (3.2).

4.1 Latent class membership estimations for LCA

Latent class analysis is a useful tool to classify objects based on their responses to a set of categorical items. Suppose the basic model has an underlying categorical latent variable $S_i = 1, \ldots, J$ for individual *i*, and within any latent class, the measured indicators are assumed to be independent of one another. Therefore, if we can estimate the unobservable class membership S_i, viewing the estimated class membership as known variable, then it is easy to predict the parameters in (3.1). We propose the following strategy to estimate the unobservable class membership S_i.

Our strategies are to apply the concept of *k-means* (MacQueen, 1967) and *divisive hierarchical methods* to cluster the objects. Here we do not cluster the objects into *J* subgroups such that the objects in one subgroup are “far from” the objects in the others; we want to group objects such that response variables are statistically independent within latent classes. So we apply sample correlation or sample covariance as distance in *k-means* and *divisive hierarchical methods* and the “loss information” and “minimum loss of information” concepts in Ward’s hierarchical clustering method.

Now we illustrate how to calculate the sample correlation and sample covariance matrix in *k-means* and *divisive hierarchical methods*.

For individual *i*, we transform the *M* polytomous outcome indicators $(Y_{i1},\dots,Y_{iM})$ to the dummy variables

$$\tilde{\mathbf{Y}}_i=\left(Y_{i11},\dots,Y_{i1(K_1-1)},Y_{i21},\dots,Y_{i2(K_2-1)},\dots,Y_{iM1},\dots,Y_{iM(K_M-1)}\right)$$

with $Y_{imk}=I(Y_{im}=k), m=1,\dots,M, k=1,\dots,K_m-1$.

and variance-covariance matrix

$$\mathrm{Cov}\left(\tilde{\mathbf{Y}}_i\right)=\left[\mathrm{Cov}\left(Y_{imk},Y_{iqs}\right)\right]=\begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1M}\\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2M}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{B}_{M1} & \mathbf{B}_{M2} & \cdots & \mathbf{B}_{MM}\end{bmatrix}, \qquad (4.1)$$

where for the *m*th item and *q*th item, $\mathbf{B}_{mq}$ is the block of $(K_m-1)\times(K_q-1)$ covariance matrix. Various elements of the variance-covariance matrix of measured indicators are

$$\mathrm{Cov}\left(Y_{imk},Y_{iqs}\right)=\begin{cases}\Pr(Y_{imk}=1)-\Pr(Y_{imk}=1)\Pr(Y_{iqs}=1) & \text{if } m=q \text{ and } k=s\\ -\Pr(Y_{imk}=1)\Pr(Y_{iqs}=1) & \text{if } m=q \text{ and } k\neq s\\ \Pr(Y_{imk}=1,Y_{iqs}=1)-\Pr(Y_{imk}=1)\Pr(Y_{iqs}=1) & \text{if } m\neq q\end{cases} \qquad (4.2)$$

These variances were estimated by using the sample averages. Furthermore, we can also calculate the sample correlation matrix as $\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathrm{Cov}\left(\tilde{\mathbf{Y}}_\mathrm{i}\right)\tilde{\mathbf{D}}^{-\frac{1}{2}}$, where $\tilde{\mathbf{D}}=\mathrm{diag}\left(\hat{\mathbf{B}}_{11},\hat{\mathbf{B}}_{22},\dots,\hat{\mathbf{B}}_{MM}\right)$. There are *k-means* and *divisive hierarchical* clustering algorithm separately.

K-means algorithm:

1. First, all objects are partitioned into *K* initial clusters.
2. Proceed through the list of objects, assigning an object to the cluster such that "minimum loss of independence" is reached.
3. Repeat step 2 until no more reassignments take place.

In step 1, we specify *K* preliminary centroids (seed points) and then proceed through the list of objects, assigning an object to the cluster whose centroid (mean) is nearest and the distance is computed using Euclidean distance. Since we use the sample covariance or correlation to measure minimum loss of independence, it is necessary to reach enough sample size in each initial cluster. Once an initial cluster including members less than we expected, we adjust the number of objects in each cluster by repartitioning the objects "randomly" and "evenly" into K initial clusters.

Now, we introduce the concept of minimum loss of independence in step 2. Denote that $\mathrm{MCov_k}$ be the mean of the absolute values of entries in non-diagonal blocks of sample correlation/covariance matrix in a given cluster *k*. For a given object, if it is assigned to some cluster *j*, we define the loss of independence $\mathrm{LoI_j}$ as the sum of the $\mathrm{MCov_k}$, that is, $\mathrm{LoI_j} = \mathrm{MCov_1^{(j)}} + \mathrm{MCov_2^{(j)}} + \cdots + \mathrm{MCov_K^{(j)}}$, where $\mathrm{MCov_k^{(j)}}$ is the mean of the absolute values of non-diagonal-block entries of correlation/covariance matrix after the object being assigned to cluster *j*. After assigning some object through *K* clusters, we can obtain $\mathrm{LoI_j}$, $j = 1, \ldots, K$. The smaller the value of $\mathrm{LoI_j}$, the more independent the observed variables for objects within cluster *j* are. Then we take the minimum $\mathrm{LoI_j}$ as the "minimum loss of independence" and assign a given object to the cluster corresponding to the minimum loss of independence. Figure 1 will display *k-means* algorithm procedure.

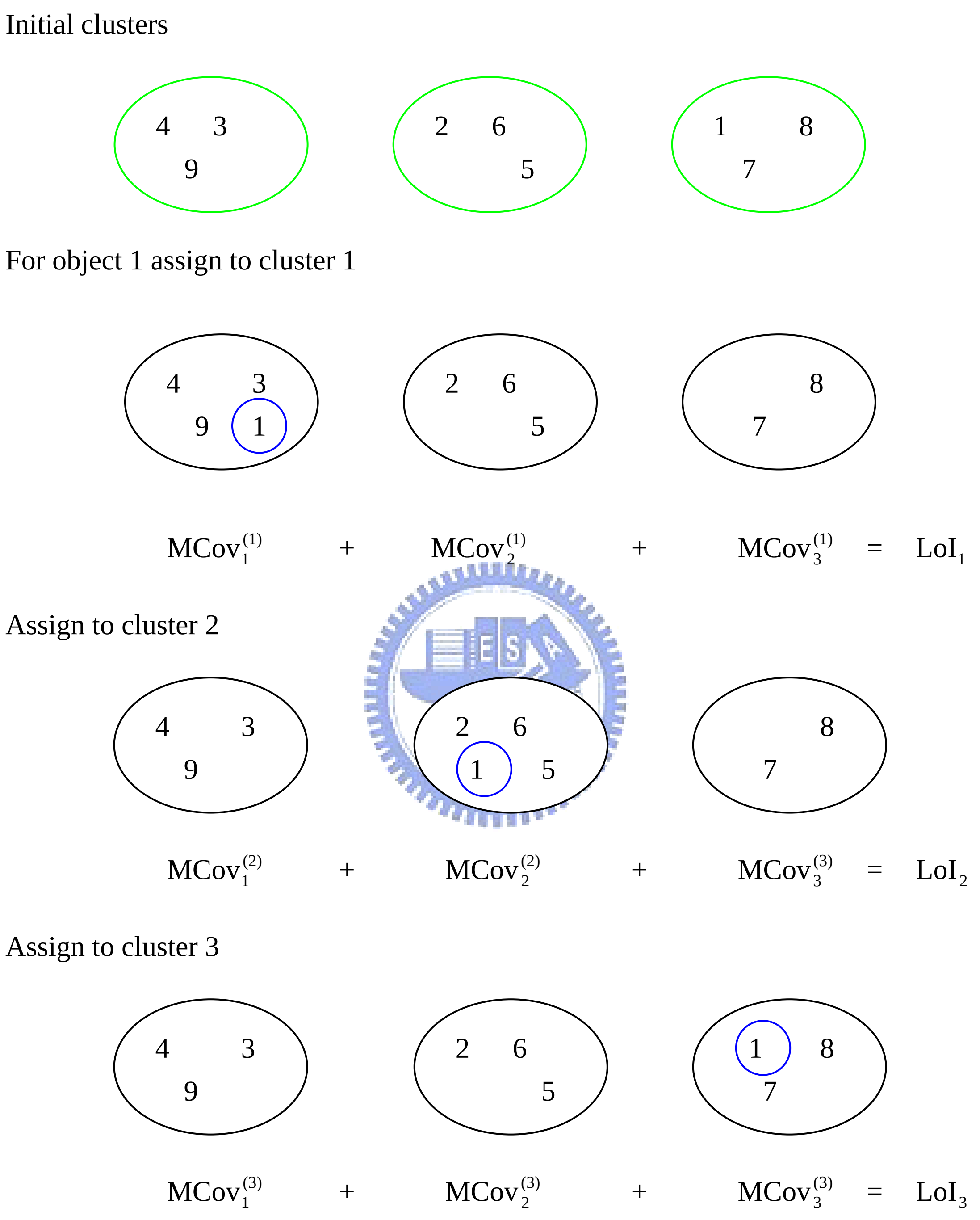

Figure 1: An example of k-means algorithm procedure.

Step 1: Partition 9 objects into 3 initial clusters.

Step 2: What cluster the objects will be assigned to?

Assigning the object 1 into cluster 1, 2 and 3 separately, and we can obtain LoI_1, LoI_2, and LoI_3. Assign the object 1 into the cluster which attaining "minimum loss of independence".

Step 3: Repeat step 2 until no more reassignments take place.

Divisive hierarchical clustering algorithm:

1. Start with a single cluster containing all objects.
2. To divide the preliminary cluster, we apply *k-means* approach above to get the two smaller clusters.
3. We divide one of two clusters such that the 'minimum loss of independence' is attained.
4. Repeat Step 3 until no more division take place.

Here we illustrate the detail in Step 3. For currently K clusters, which one cluster we divided first? We divide cluster such that the minimum loss of independence is reached. For a given cluster j, if it is divided into two smaller clusters, U and V. We define the loss of independence LoI_j as the sum of $MCov_k$ (defined in K-means algorithm) of each cluster.

$$LoI_j = MCov_1 + \cdots + MCov_{(j-1)} + MCov_U + MCov_V + MCov_{(j+1)} + \cdots + MCov_K .$$

The smaller the value of LoI_j is, the more independent the observed variables for objects within cluster U and V are. So, we take the minimum LoI_j as the "minimum loss of independence" and divide the cluster j whose division results in the minimum loss of independence. An example of divisive hierarchical algorithm procedure can be found in Figure 2.

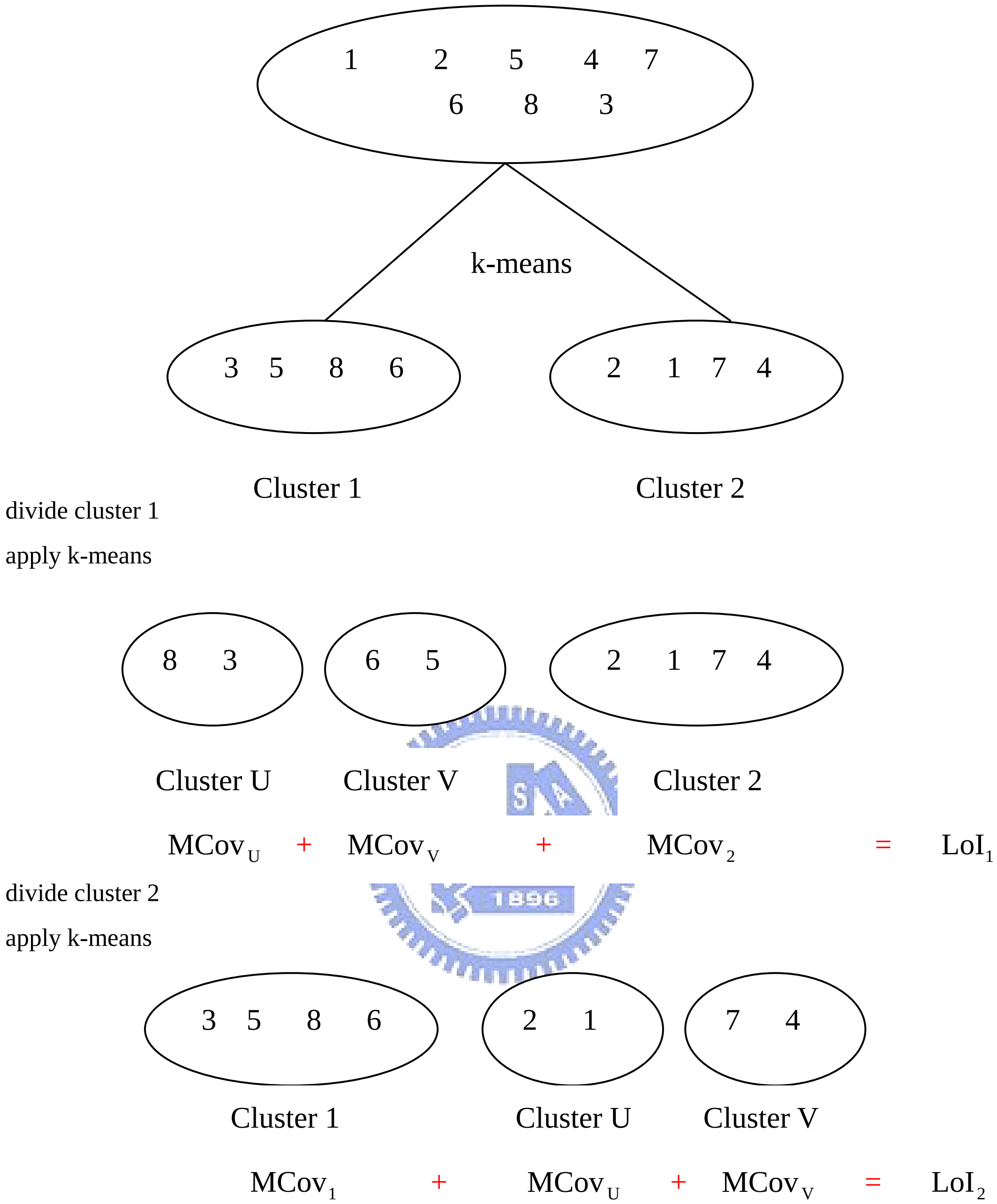

Figure 2: An example of divisive hierarchical algorithm procedure.

Step 1: Start with a single cluster which consists of all objects.

Step 2: Using k-means approach to divide the initial cluster into two smaller clusters.

Step 3: Which cluster will be divided first?

Consider the divisions of all current clusters, we get LoI_1 and LoI_2. Divide the cluster whose division results in the “minimum loss of independence”.

Step 4: Repeat Step 3 until no more division take place.

The results of divisive hierarchical clustering method can be displayed as a dendrogram. The vertical axis gives the values of one minus minimum loss of independence at which the division occurs.

4.2 Latent class membership estimations for RLCA

The *k-means* and *divisive hierarchical methods* we proposed also work for the model (3.2) under eliminating the covariate effects (Huang, 2005) and "marginalize" the model (3.2).

The key to marginalizing over $\mathbf{z}_i$ is that the process must yield random variables that follow a finite mixture distribution that is both independent of $\mathbf{z}_i$ and has J mixing components. One strategy for achieving such marginalization can be motivated by the properties of added variable plots for linear regression models. The conditional probabilities (3.4) can be viewed as fitting a logistic regression of $\mathbf{Y}_{im}$ on S_{i} adjusting for $\mathbf{z}_{im}$, separately for each m. Then the problem becomes how to calculate residuals from the generalized linear model:

$$\mathrm{logit}\left[\mathrm{E}\left(Y_{im} \mid \mathbf{S}_i, \mathbf{Z}_{im}^c\right)\right] = \left(\mathbf{Z}_{im}^c\right)^{\mathrm{T}} \boldsymbol{\alpha}_m^* \quad \text{for} \quad i = 1,\ldots,N\,;\, m = 1,\ldots,M \tag{4.3}$$

where "p" denotes polytomous responses; $\mathbf{Y}_{im}^p = \left[Y_{im1},\ldots,Y_{im(K_m\text{-}1)}\right]^{\mathrm{T}}$ and $Y_{imk} = I\left(Y_{im} = k\right)$; $\mathbf{Z}_{im}^c = \left[\left(z_{im1} - \bar{z}_{m1}\right),\ldots,\left(z_{imL} - \bar{z}_{mL}\right)\right]^{\mathrm{T}}$, ("centered" covariate vector); $\bar{z}_{mp} = \left(1/N\right)\sum_{i=1}^{N} z_{imp}$;

Under polytomous item responses, the pseudo-residual of ith participant's mth response item is

$$\mathbf{R}_{im}^p = \left(\hat{\mathbf{V}}_{im}^p\right)^{-1}\left(\mathbf{Y}_{im}^p - \hat{\boldsymbol{\mu}}_{im}^p\right) \tag{4.4}$$

where “hat” denotes the estimated values; $\mathbf{R}_{im}^{p}=\left[R_{im1},\ldots,R_{im(K_m\text{-}1)}\right]^{\mathrm{T}}$; $\mathbf{V}_{im}^{p}=\mathrm{Var}\left(\mathbf{Y}_{im}^{p}\right)$; $\boldsymbol{\mu}_{m}^{p}=\mathrm{E}\left(\mathbf{Y}_{m}^{p}|\mathbf{Z}_{m}^{c}\right)$; and $i=1,\ldots,N$; $m=1,\ldots,M$; $k=1,\ldots,K_m$.

When eliminating $\mathbf{z}_i$, we have the nice property that the covariates associated with class prevalences $\mathbf{x}_i$ can be ignored and under the assumption of $\mathbf{x}_i$ and $\mathbf{z}_{im}$ are independent, we can treating the residuals from the model (4.1) as new response variables. Details of the above the marginalization process can be found in Huang (2005) and in section 2.3 of this thesis. We can classify objects based on the new response variables $\mathbf{R}_{im}^{p}$ to a set of categorical items. The methods to classify objects are the same as the k-means and divisive hierarchical clustering algorithms in section 4.1, besides the estimation of the covariance matrix $\mathrm{Cov}\left(\tilde{\mathbf{Y}}_i\right)$ in (4.1), evaluated as $\frac{1}{n\text{-}1}\left[\tilde{\mathbf{R}}^{\mathrm{T}}\left(\mathbf{I}_{\mathrm{n}}\text{-}\frac{1}{n}\mathbf{1}\mathbf{1}'\right)\tilde{\mathbf{R}}\right]$, where $\tilde{\mathbf{R}}$ is the residual matrix of n objects.

4.3 Parameter estimation by viewing estimated latent class as known variable

When using *k-means* and *divisive hierarchical method* to estimate the latent class membership, we denote the estimated latent class as $\hat{S}_{\mathrm{i}}$ for individual *i*. Replace S_{i} by $\hat{S}_{\mathrm{i}}$ in (3.3) and (3.4) as the following:

$$\log\left[\frac{\Pr\left(\hat{S}_{\mathrm{i}}=\mathrm{j}\right)}{\Pr\left(\hat{S}_{\mathrm{i}}=\mathrm{J}\right)}\right]=\beta_{0j}+\beta_{1j}x_{i1}+\cdots+\beta_{pj}x_{ip} \tag{4.5}$$

and

$$\log\left[\frac{\Pr\left(Y_{im}=k|\hat{S}_i,\mathbf{z}_i\right)}{\Pr\left(Y_{im}=K|\hat{S}_i,\mathbf{z}_i\right)}\right]=\gamma^{*}_{mk0}+\gamma^{*}_{mk1}\hat{S}_{i1}+\cdots+\gamma^{*}_{mk(\mathrm{J}\text{-}1)}\hat{S}_{i(J\text{-}1)}+\alpha_{1mk}z_{im1}+\cdots+\alpha_{Lmk}z_{imL} \tag{4.6}$$

$$i=1,\ldots,N;\ m=1,\ldots,M;\ k=1,\ldots,(K_m-1);\ j=1,\ldots,(J\text{-}1)$$

where $\hat{S}_{ij}=I\left(\hat{S}_i = j\right)$ and $\gamma^{*}_{mk0} = \gamma_{mkJ}$, $\gamma^{*}_{mkj} = \gamma_{mkj} - \gamma_{mkJ}$ in (3.4).

Then, it is easily to estimate the parameters in (3.2) using multinomial logistic regression (4.5) and (4.6).

5. Simulation study

The simulation study aims to evaluate the performance of the proposed approach.

5.1 Generated data from RLCA model

Three different RLCA models are simulated in our simulation study. The first was a three-class RLCA with five two-level measured indicators, two covariates associated with conditional probabilities, and two covariates associated with latent prevalences (i.e., $J=3, M=5, K_1=\cdots=K_5=2, P=L=2$). The second was a six-class RLCA with five three-level measured indicators and the same setting as the three-class model (i.e., $J=6, M=5, K_1=\cdots=K_5=3, P=L=2$). The last was a two-class RLCA with five three-level measured indicators, two covariates with conditional probabilities and two covariates latent prevalences (i.e., $J=2, M=5, K_1=\cdots=K_5=3, P=L=2$). For each model, the model parameters $\{\beta_{pj}, \mathrm{j}=1,\cdots,\mathrm{J}-1\}$ for each $p\in\{0,1,\cdots,\mathrm{P}\}$, $\{\gamma_{jmk}, \mathrm{j}=1,\cdots,\mathrm{J}\}$ for all m, k, and $\{\alpha_{qmk}, \mathrm{m}=1,\cdots,\mathrm{M}; \mathrm{k}=1,\cdots,(\mathrm{K_m}-1)\}$ for all q, were given. Table 1~6 shows the values of parameters for the three models separately.

The covariates of three-class model, we got from the subjects who joined the Multidimensional Psychopathological Study on Schizophrenia (MPSS) or the Study on Etiological Factors of Schizophrenia (SEFOS). We got the covariates of six-class and two class models from the subjects who joined the Multidimensional Psychopathology Group Research Projects (MPGRP), MPSS or SEFOS. In three models, the covariates associated with conditional probabilities include variables of sex and age (year), and the covariates associated with latent prevalences include variables of occupation (with versus without occupation) and dprime, which is the sensitivity index of the Continuous Performance Task (CPT; Rosvold et al., 1956) performance.

We fit each model under several different sample sizes. For the three-class RLCA, the selected sample sizes were 100 and 500, which gave roughly 3 and 16 individuals per parameter of RLCA (3.2), respectively. For the six-class RLCA, the selected sample sizes were 300 and 1000, which gave roughly 3 and 10 individuals per parameter, respectively. For the two-class RLCA, the selected sample sizes were 150 and 700, which gave roughly 3 and 16 individuals per parameter, respectively. The observable measurements $\mathbf{Y}_i$ were then generated from each different model structure with 100 replications.

5.2 Simulation results

In each case, the results of simulation study are represented in six tables which include the average parameters estimates, average conditional probabilities, average latent prevalences, average correlation coefficients, and average match proportions for 100 replications separately. We shall explain these results later. The simulation results for 3-class model with 100 sample sizes are presented from Table 7 to Table 12. The simulation results for 3-class model with 500 sample sizes are presented from Table 13 to Table 18. The simulation results for 6-class model with 300 sample sizes are presented from Table 19 to Table 24. The simulation results for 6-class model with 1000 sample sizes are presented from Table 25 to Table 30. The simulation results for 2-class model with 150 sample sizes are presented from Table 31 to Table 36. The simulation results for 2-class model with 700 sample sizes are presented from Table 36 to Table 42. According to Table 7 ~ Table 42, we can see that these results of the k-means method and divisive hierarchical using correlation coefficients measurement are similar to those of k-means method and divisive hierarchical using covariance measurement. So, we shall discard the results of k-means method and divisive hierarchical using covariance measurement in the following discussion.

First we discuss the simulation results which are presented from Table 13 to Table 18 of 3-class model with 500 sample sizes.

Average parameters estimations

Table 12 and Table 13 under the column "TRUE" include all $\{\beta_{pj}, \gamma_{jmk}, \alpha_{qmk}\}$ in simulated data. All average of $\{\beta_{pj}, \gamma_{jmk}, \alpha_{qmk}\}$ estimates got from the k-means method using correlation coefficient measurement (K_Corr) and covariance measurement (K_Cova) separately and the divisive hierarchical method using correlation coefficient measurement (D_Corr) and covariance measurement (D_Cova) appeared in Table 12 and Table 13. Table 12 and Table 13 can demonstrate that the parameters estimates got from the k-means method are well compared to the true parameters. But the parameters estimates got from the divisive hierarchical procedure are poor. Furthermore, the divisive hierarchical procedure is sensitive to cluster structure. This means that hierarchical procedure have the chance to perform more well only when there is clear cluster structure than when there is no clear cluster structure.

Table 12 and Table 13 also include the standard errors of parameters estimates in doing multinomial regressions, (4.1) and (4.2), and the average sample standard errors of the parameters estimates for 100 replications. The sample standard errors of the estimates for 100 replications include the variation of doing multinomial regression and creating cluster membership. Because we use the multinomial regression to estimate parameters under the assumption of known cluster membership, the standard errors of parameters estimates in doing multinomial regression did not include the variations of creating cluster membership. Therefore, the standard errors of parameters estimates in doing multinomial regressions should be smaller than the sample standard errors of the estimates for 100 replications. This is demonstrated in

Table 12 and Table 13. However this is not demonstrated in Table 7 and Table 8 for the 3-class model with 100 sample sizes which gave few individuals per parameter. For the sparse data, the estimated standard errors of parameters estimates in doing multinomial regressions are not accurate. Therefore, the standard errors of parameters estimates in doing multinomial regressions are not always smaller than the sample standard errors of the estimates over 100 replications for the 3-class model with 100 sample sizes.

Average Conditional Probabilities

Table 14 under the column "TRUE" displays the RLCA conditional probabilities evaluated at the sample means of the incorporated covariates:

$$\mathrm{p}_{\mathrm{mkj}} = \frac{\exp(\gamma_{\mathrm{mkj}} + \bar{z}_m^T \alpha_{mk})}{1 + \sum_{i=1}^{K-1} \exp(\gamma_{\mathrm{mkj}} + \bar{z}_m^T \alpha_{mk})} \quad , \quad \mathrm{k} = 1, \cdots, \mathrm{K} - 1$$

$$\mathrm{p}_{\mathrm{mKj}} = \frac{1}{1 + \sum_{i=1}^{K-1} \exp(\gamma_{\mathrm{mkj}} + \bar{z}_m^T \alpha_{mk})}$$

where $\bar{\mathbf{z}}_m = \frac{1}{N} \sum_{i=1}^{N} \mathbf{z}_{im}$.

The average of estimated conditional probabilities over 100 replications with k-means and divisive hierarchical methods appear in Table 14. The estimated conditional probabilities for k-means and divisive hierarchical methods are

$$\hat{p}_{mkj} = \frac{\text{the number of individuals in class } j \text{ being at level } k \text{ of } Y_{im}}{\text{the number of individuals in class } j}$$

Overall, the average conditional probabilities for the k-means method are more closed to the true conditional probabilities than the average conditional probabilities for the divisive hierarchical method.

Average Latent Prevalence:

Table 15 under the column "TRUE" displays the sample average of the RLCA prevalences:

$$\eta_j^* = \frac{1}{\mathrm{N}} \sum_{\mathrm{i=1}}^{\mathrm{N}} \eta_j (\mathrm{x_i^T} \beta)$$

The average of estimated prevalences over 100 replications with k-means and divisive hierarchical methods are also shown in Table 16. The estimated prevalences are

$$\hat{\eta}_j = \frac{\text{the number of individuals in class } j}{\text{the total number of individuals in study}}$$

Overall, the average latent prevalences for the k-means method are more closed to the true prevalences than the average latent prevalence for the divisive hierarchical clustering method.

Average Correlation Coefficients

We evaluated the $\mathrm{MCov_k}$ of the objects in the same cluster *k*. Table 16 displays the average of $\mathrm{MCov_k}$ over 100 replications in each cluster *k*. The k-means approach resulted smaller average correlation coefficients than the divisive hierarchical method.

Next, for the 6-class model with 1000 sample sizes, we shall discuss the simulation results which are presented from Table 22 to Table 26. These tables show that the results of whether the k-means procedure or the divisive hierarchical procedure are poor obviously comparing to the 3-class model with 500 sample sizes. The 2-class model with 150 and 700 sample sizes, we shall discuss the simulation results which are presented from Table 27 to Table 36. It is reasonable that the results of k-means and divisive hierarchical clustering methods for 2-class RLCA models are the same.

When we use maximum likelihood to estimate the parameters in (3.2), the maximum likelihood estimation (MLE) is relative to the number of individuals given in per parameter. For the spare data which gave less individuals per parameter, the MLE can not be obtained or the MLE is not a good estimation .For the three models, 3-class RLCA with 100 sample sizes, 6-class RLCA with 300 sample sizes and 2-class RLCA with 150 sample sizes, which gave less individuals per parameter, the simulation results are not wore than those that gave more individuals per parameter. It demonstrates that our clustering procedure is irrelative to the number of individuals given per parameter.

6. Example

Schizophrenia Data

The present study is composed of three projects, the Multidimensional Psychopathology Group Research Projects (MPGRP), the Multidimensional Psychopathological Study on Schizophrenia (MPSS) and the Study on Etiological Factors of Schizophrenia (SEFOS). The initial project MPGRP investigated the clinical manifestations of schizophrenia in a cohort of schizophrenia patients. The subsequent project MPSS focused on the follow-up neuropsychological evaluation of the MPGRP patients. The project SEFOS aimed to search for neurobiological, environmental and genetic factors underlying schizophrenia. The analyzed data include 169 acute-state patients who had completed the PANSS within one week of index admission and 161 subsided-state patients who were living with community and under family care.

The major instrument applied in this study is the PANSS, were used to collect patients' symptom measurements, an assessment of the clinical psychopathological symptoms of schizophrenia. It has 30 items rated on a 7-point scale (1=absent, 7=extreme). The PANSS consists of three subscales: positive (seven symptoms: P1-P7), negative (seven symptoms: N1-N7), and general psychopathology (sixteen symptoms: G1-G16). Because the original 7-point scale is too complex and has too many parameters to analyze, we reduced the 7-point scale on PANSS by merging the scales which have the percentages less than 5% on each item.

Demographic variables included gender, age, onset-age of psychotic symptoms, years of education, and occupation (having versus no occupation). The category of no occupation included housewives, students, unemployed and retired people.

The environmental factors were related to obstetric complications, prenatal growth retardation, special personal behavior and psychological adjustment problems. There were three environmental questions including: (1) the patient had brain injury in the developmental process, such as premature birth, brain damage and retarded intelligence; (2) the patient had unstable mood or abnormal behavioral traits to interfere with daily life, including angry, timid, depressed and inactive; and (3) the patient had psychological adjustment problems to interfere with daily life, including bad relation between parents, getting along badly with sibling, getting physical disease and unforeseen happenings of family. All three environmental factors were rated by a 3-point scale with 0 as no event, 1 as slight and no obvious effect on emotional and behavioral reacting, and 2 as obvious effect on emotional and behavioral reacting.

The neuropsychological batteries assessed reaction time, attention, speed of information processing, and active problem solving. Specifically, the test batteries included several standard neuropsychological instruments with demonstrated reliability and validity, including the Continuous Performance Test (CPT), Wisconsin Card Sorting Test (WCST), Wechsler Adult Intelligence Scale-Revised (WAIS-R), Wechsler Memory Scale-Revised (WMS-R), and Trail Making Tests A and B (TMT-A and -B). Here we concentrated on CPT.

We fit RLCA model with 30 7-level measured indicators, the covariates associated with conditional probabilities include variables of sex, age (year), years of education (year), and occupation (with versus without occupation), and the covariates associated with latent prevalences include variables of age of onset (year), envir11, envir21, envir22, envir31, envir32, and dprime.

We group objects by k-means and divisive hierarchical approaches, and the analysis reported here aims to describe the associations between risk factors and underlying latent class, and examine the composition of patient subtypes across different disease states.

Here, we introduced a useful tool for clustering. Heatmap has the notion of rearranging the columns and rows to show structure in the data. A heatmap is a two-dimensional, rectangular, colored grid. It displays data that themselves come in the form of a rectangular matrix. The color of each rectangle is determined by the value of the corresponding entry in the matrix. The rows and columns of the matrix can be rearranged independently. Usually they are reordered so that similar rows are placed next to each other, and the same for columns. Among the orderings that are widely used are those derived from a hierarchical clustering, but many other orderings are possible. If hierarchical clustering is used, then it is customary that the dendrograms are provided as well. In many cases the resulting image has rectangular regions that are relatively homogeneous and hence the graphic can aid in determining which rows (generally the genes) have similar expression values within which subgroups of samples (generally the columns).

Results for patients at the acute state by divisive hierarchical clustering method

Heatmap for patients at the acute state was shown in Figure 3. The column dendrogram is agglomerative hierarchical clustering method with distance measurement using one minus correlation and the row dendrogram is our divisive hierarchical clustering with distance measurement using one minus loss of independence. The color of each cell represented the extent of induction or repression of a given gene.

Although the heatmap did not display the class structure clearly, we can use the dendrogram of divisive hierarchical method at the left to group objects into four classes.

Table (37) contains the scores (mean ± standard error) of 30 items (or 5 factors) in each class, we can characterize four classes as follows. Class 1 has lower scores (mean) on the factor 2, factor 3, and factor 4. Class 3 has higher scores (mean) on the factor 2, factor 3, and factor 5. The scores of four classes on the factor 1 are similar.

Table (38) includes odds ratios for the relationship between classes of schizophrenia at the acute state and demographic/environmental/neuropsychological variables. Notice that odds ratios for each class were compared with the class 4 (the reference class).

By comparing with the patients of the class 4, patients of the class 1 were more likely to be female. Patients of the class 2 tended to have unstable mood or abnormal behavior to interfere (obviously) with their life.

Results for patients at the subside state by divisive hierarchical clustering method

Heatmap for patients at the subsided state was shown in Figure 4. We clustered the objects into three classes.

Table (39) includes the scores (mean ± standard error) of 30 items (or 4 factors) in each class, we can characterize three classes as follows. Class 1 has higher scores (mean) on each factor. The scores in class 2 and class 3 are not different on each factor.

Table (40) contains the odds ratios for the relationship between class of the subsided schizophrenia and demographic/environmental/neuropsychological variables. Odds ratios for each subtype are compared with the class 3 (the reference subtype). By comparing with the patients of the class 3, patients of the class 1 tended to be younger age of onset and have unstable mood or abnormal behavior to interfere (obvious) with their life. Patients of the class 2 were more likely to have psychological adjustment problems to slightly interfere with their life.

Results for patients at the acute state by k-means clustering method

Table (41) includes the scores (mean ± standard error) of 30 items (or 5 factors) in each class. Class 3 has higher scores on each factor and class 2 has lower scores on factor 2, factor 3, factor 4, and factor 5. However, class 1 in Table (37) may correspond to class 2 in Table (41). These results clustered by k-means method are closed to the results clustered by divisive hierarchical method.

Table (42) contains the odds ratios for the relationship between class of the acute schizophrenia and demographic/environmental/neuropsychological variables. Odds ratios for each subtype are compared with the class 4 (the reference subtype). By comparing with the patients of the class 4, patients of the class 1 and class 2 were more likely to be female. The results of k-means clustering method are closed to the results of divisive hierarchical method.

Results for patients at the subside state by k-means clustering method

Table (43) includes the scores (mean ± standard error) of 30 items (or 4 factors) in each class. Class 1 has higher scores (mean) and the scores in class 3 are lower on each factor.

Table (44) contains the odds ratios for the relationship between class of the subsided schizophrenia and demographic/environmental/neuropsychological variables. Odds ratios for each subtype are compared with the class 3 (the reference subtype). By comparing with the patients of the class 3, patients of the class 1 and class tended to be younger age of onset.

Breast Cancer Data

Here we used DNA microarray analysis on primary breast tumours of 117 young patients. The 78 sporadic lymph-node-negative patients under 55 years of age were selected specifically to search for a prognostic signature in their gene expression profiles. Forty-four patients remained free of disease after their initial diagnosis for an interval of at least 5 years (good prognosis group, mean follow-up of 8.7 years), and 34 patients had developed distant metastases within 5 years (poor prognosis group, mean time to metastases 2.5 years). This dataset record the mean ratio of the intensities of the red and green channels, this reflects the extent of induction or repression of a given gene, and p-values, means that a gene's mean ratio is significantly different from 1, or no change. Besides, the covariates, age (year) and metastasis of year (1, if metastases > 5 years; 0, otherwise), are contained.

This gene expression microarray experiments can generate data sets with multiple missing expression values. Many algorithms for gene expression analysis require a complete matrix of gene array values as input. For example, methods such as hierarchical clustering and *k-means* clustering are not robust to missing data, and may lose effectiveness even with a few missing values. Methods for imputing missing data are needed, therefore, to minimize the effect of incomplete data sets on analyses, and to increase the range of data sets to which these algorithms can be applied. Troyanskaya et al. (2001) suggests that k-nearest neighbors (KNN) approach provides accurate and robust estimates of missing values. The KNN-based method selects genes with expression profiles similar to the gene of interest to impute missing values. For instance, if we consider gene A that has one mission value in experiment 1, this method would find K other genes, which have a value present in experiment 1 , with expression most similar to A in experiments 2-N (where N is the total number of experiments). A weighted average of values in experiment 1 from the K closest genes is then used as an estimate for the missing value in gene A. In the weighted average, the contribution of each gene is weighted by similarity of its expression to that of gene A.

In brief, approximately 5,000 genes (with at least a twofold difference and a p-value of less than 0.01 in more than five tumours) were selected from the 25,000 genes.

Standardizing the data in this fashion achieves a location and scale normalization of the different arrays. In a study of normalization methods, we have found scale adjustment to be desirable in some cases to prevent the expression levels in one particular array (Yang et al. 2001). These 5,000 genes were standardized so that the observations (arrays) have mean 0 and variance 1 across variables (genes).

Many genes exhibit near-constant expression levels across tumor samples. We thus applied a preliminary selection of genes based on the ratio of their between-group to within-group sums of squares. For a gene j, this ratio is

$$BW(j)=\frac{\sum_i\sum_k I(y_i=k)(\bar{x}_{kj}-\bar{x}_{.j})^2}{\sum_i\sum_k I(y_i=k)(x_{ij}-\bar{x}_{kj})^2}, \tag{6.1}$$

where $\bar{x}_{.j}$ and $\bar{x}_{kj}$ denote the average expression level of gene j across all tumor samples and across samples belonging to class k only.

We use (6.1) to compute BW ratio for each gene and select 70 genes with larger BW ratios from 5,000 genes for our study.

For the continuous data, RLCA model, with 70 measured indicators, the covariates associated with conditional probabilities include variables of age (year), and the covariates associated with latent prevalences include variables of metastasis of year, can be applied when rewriting (3.2) as the following

$$f(\mathbf{y}_i \mid \mathbf{x}_i, \mathbf{z}_i)=\sum_{j=1}^{J}\left\{\Pr(S_i=j \mid \mathbf{x}_i)\prod_{m=1}^{M} f(\mathrm{y}_{im} \mid S_i=j, \mathbf{z}_{im})\right\},$$

where $f(y_{im} \mid S_i = j, \mathbf{z}_{im}) \sim N\left(\mu_{imj}, \sigma_m^2\right)$. The parameters μ_{imj} and σ_m^2 can be replaced by the estimations as the following

$$\hat{\mu}_{imj} = \mathrm{E}\left(y_{im} \mid S_i = j, \mathbf{z}_{im}\right) = \hat{\gamma}_{m0} + \hat{\gamma}_{m1} S_{i1} + \cdots + \hat{\gamma}_{m(J-1)} S_{i(J-1)} + \hat{\alpha}_{1m} z_{im1} + \cdots + \hat{\alpha}_{Lm} z_{imL}$$

$\hat{\sigma}_m^2$: the residual mean square of m linear regression model .

It can also predict the class membership of the additional data using the posterior probability of class membership

$$\begin{aligned}\theta_{ij} &= \Pr\left(S_i = j \mid \mathbf{y}_i, \mathbf{x}_i, \mathbf{z}_i\right) \\ &= \frac{\Pr\left(S_i = j \mid \mathbf{x}_i\right) \prod_{m=1}^{M} f\left(\mathrm{y}_{im} \mid S_i = j, \mathbf{z}_{im}\right)}{\sum_{l=1}^{J} \left\{\Pr\left(S_i = l \mid \mathbf{x}_i\right) \prod_{m=1}^{M} f\left(\mathrm{y}_{im} \mid S_i = l, \mathbf{z}_{im}\right)\right\}}\end{aligned}$$

$$i = 1, \ldots, N \;;\; m = 1, \ldots, M \;;\; j = 1, \ldots, (J\text{-}1)$$

Results for breast cancer data with divisive hierarchical clustering method

Heatmap was applied to microarray data by Eisen et al. (1998) and have become a standard visualization method for this type of data.

The heatmap for 70-gene profile is displayed in Figure 5. The column dendrogram is agglomerative hierarchical clustering method with distance measurement using one minus correlation and the row dendrogram is our divisive hierarchical clustering with distance measurement using one minus loss of independence. The color of each cell represented the extent of induction or repression of a given gene. We can easily group objects into two classes and include 39 objects in each class. Notably, in the upper group only 30% of the patients were from the group who developed distant metastases greater than 5 years, whereas in the lower group 82% of the patients had good prognosis disease. Thus we can distinguish between “good prognosis” and “poor prognosis” patients.

We fit a two-class RLCA model with the covariates associated with conditional probabilities include variable of age (year) and the covariates associated with latent prevalences include variable of metastasis more than 5 years (1, if more than 5 years; 0, otherwise).

Table (45) includes odds ratios for the relationship between classes of tumours and age and variable of metastasis more than 5 years. Notice that odds ratios for each class were compared with the class 2. By comparing with the patients of the class 2, patients of the class 1 were more likely to be metastasis more than 5 years.

To validate our method, an additional independent set of primary tumours from 19 young, lymph-node-negative breast cancer patients was selected. This group consisted of 7 patients who remained metastasis free for at least five years, and 12 patients who developed distant metastases within five years. The disease outcome was predicted by the posterior probability of class membership in Table (46) and resulted in 3 out of 19 incorrect classifications.

Results for breast cancer data with k-means clustering method

Table (47) includes odds ratios for the relationship between classes of tumours and age and variable of metastasis more than 5 years, and Table (48) displayed that the performance of predictions of class membership. These results consist with the results of divisive hierarchical method, because the k-means method is a case of divisive hierarchical clustering method for grouping 2-class.

7. Discussion

One has demonstrated that the agglomerative hierarchical approach didn't perform well in the same simulation studies before. Since each cluster contains only very few objects at the early stage of the agglomerative hierarchical method, it is not appropriate to use covariance and correlation to measure the independence between two items. The wrong reallocation of objects at the early stage will result in wrong reallocation of objects at the following stage.

The divisive hierarchical clustering method we proposed combined the k-means and the (original) divisive hierarchical clustering ideas to improve the problem that the few objects at the early stage of agglomerative hierarchical method.

In the simulation studies, however, the results of the k-means and divisive hierarchical clustering methods have different performance in estimating parameters in RLCA. The parameters estimated by k-means approach are closed to the true parameters but the divisive hierarchical method does not perform well unexpectedly.

In the example studies, the divisive hierarchical method does the successful division especially in microarray data, and makes a good prediction for class membership.

References

Agresti, A.(1984). Analysis of Categorical Data. New York: John Wiley and Sons.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T. Jr, J.H., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P.O., and staudt, L.M. (2000), Different Types of Diffuse Large B-Cell Lymphoma Identified by Gene Expression Profiling, Nature, 403, 503-511.

Bandeen-Roche, K., Huang, G.H., Mubnoz, B., & Rubin, G.S. (1999). Determination of risk factor associations with questionnaire outcomes: A methods case study. American Journal of Epidemiology, 150, 1165-1178.

Bandeen-Roche, K., Miglioretti, D.L., Zeger, S.L., & Rathouz, P.J. (1997). Latent variable regression for multiple discrete outcomes. Journal of the American Statistical Association, 92, 1375-1386.

Cook, R.D., &Weisberg, S. (1982). Residuals and Influence in Regression. London: Chapman Hall.

Dayton, C.M., &Macready, G.B. (1988). Concomitant-variable latent-class models. Journal of the American Statistical Association, 83, 173-178.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1997). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Toyal Statistical Society, Series B, 39, 1-38.

Dudoit, S., Fridlyand, J., & Speed, T.P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of the American Statistical Association, 97, 77-87.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, 61, 215-231.

Green, B.F. (1951). A general solution of the latent class model of latent structure analysis and latent profile analysis. Psychometrika, 16, 151-166.

Huang, G.H., & Bandeen-Roche, K. (2004). Building an identifiable latent variable model with covariate effects on underlying and measured variables. Psychometrika, 69, 5-32.

Huang, G.H. (2005). Selecting the number of classes under latent class regression: A factor analytic analogue. Psychometrika, 70, 325-345.

Lazarsfeld, P.F., & Henry, N.W. (1968). Latent Structure Analysis. New York: Houghton-Mifflin.

L.J. van't Veer et al. Gene expression profiling predicts clinical outcome of breast cancer.Nature, 415, 530-536, 2002.

Macnaughton-Smith, P., Williams, W. T., Dale, M. B., and Mockett, L.G. (1964), Dissimilarity analysis: A new technique of hierarchical sub-division, Nature, 202, 1034-1035. [6.1, 6.4.1, 6.5.1, 7.5.3]

MacQueen, J.B. (1967). “Some Methods for Classification and Analysis of Multivariate Observations.” Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1, Berkeley, CA: University of California Press, 281-297.

McCullagh, P., & Nelder, J.A. (1989). Generalized Linear Models, 2th edition. London: Chapman and Hall.

Melton, B., Liang, KY., & Pulver, A.E. (1994). Extended latent class approach to the study of familial/sporadic forms of a disease: Its application to the study of the heterogeneity of schizophrenia. Genetic Epidemiology, 11, 311-327.

Morgan, Byron J.T., and Andrew P.G. Ray. (1995). “Non-uniqueness and Inversions in Cluster Analysis.” Applied Statistics, 44, 117-134.

Perou, C.M, Jeffrey, S.S., van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C.F., Lashkari, D., Shalon, D., Brown, P.O., and Botstein, D. (1999), Distinctive Gene Expression Patterns in Human Mammaru Epithelial Cells and Breast Cancers, Proceedings of the National Academy of Science, 96, 9212-9217.

Ross, D.T., Scherf, U.,Eisen, M.B., Perou, C.M., Spellman, P., Iyer, V., Jeffrey, S.S., de Rijn, M.V., Waltham, M., Pergamenschikov, A., Lee, J.C.F, Lashkari, D., Shalon, D., Myers, T.G., Weinstein, J.N., Botstein, D., and Brown, P.O., (2000), Systematic Variation in Gene Expression Patterns in Human Cancer Cell Lines, Nature Genetics, 24, 227-234.

Rosvold, H.E., Mirsk, A.F., Sarason, I., Bransome Jr, D.D. Bech, L.H. (1956). A continuous performance test of brain damage. J. Consult. Psychol., 20, 343-350.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001). Missing Value Estimation Methods for DNA Microarrays, Bioinformatics, 17, 520-525.

Tryon, R.C. (1939). Cluster Analysis. Edwards Brothers.

Van der Heijden, P.G.M., Kessens, J. & Bokenholt, U. (1996). Estimating the concomitant-variable latent-class model with the EM algorithm, Journal of Educational and Behavioral Statistics, 21, 215-229.

Ward, Jr. & Joe, H. (1963). “Hierarchical Grouping to Optimize an Objective Function.” Journal of the American Statistical Association, 58, 236-244.

Yang, Y.H., Dudoit, S., Luu, P., and Speed, T.P. (2001), Normalization for cDNA Microarray Data, in Microarrays: Optical Technologies and Informatics, eds. M.L. Bittner, Y. Chen, A. N. Dorsel, and E.R. Dougherty.

Table 1: Values of α_0 and α_{Lm} in 3 class case

	item 1	item 2	item 3	item 4	item 5
			α_0		
class 1	6.7479	3.4660	0.2348	-7.6250	11.8628
class 2	-1.3035	1.6160	-2.1880	-7.6992	-0.4100
class 3	-4.3977	-1.4305	-4.0995	-20.6690	-1.1043
			α_{Lm}		
z_{1m}	0.3944	-0.6396	-0.0581	1.0860	-0.8558
z_{2m}	0.0749	-0.0528	0.0289	0.3551	0.0189

Table 2: Values of β_0 and β_{pj} in 3 class case

	class 1 vs. class 3	class 2 vs. class 3
	β_0	
	-0.107	-0.716
	β_{pj}	
x_{i1}	-0.294	0.152
x_{i2}	0.035	-0.041

Table 3: Values of α_0 and α_{Lm} in 6 class case

	item 1	item 2	item 3	item 4	item 5
			level 1 vs. level 3		
			α_0		
class 1	5.8950	4.6522	1.0752	4.0627	3.4993
class 2	-0.1017	-0.3133	-4.4863	0.8992	0.3045
class 3	1.9543	2.8491	-2.6076	2.2020	1.9947
class 4	0.8544	0.4180	0.4860	4.7660	0.5688
class 5	-5.6842	-9.5250	-3.8248	-3.9258	-1.8959
class 6	-6.0287	-2.2211	-7.4452	-6.8831	0.1988
			α_{Lm}		
z_{1m}	-0.8438	-0.5129	-0.4503	-1.8826	-0.7673
z_{2m}	0.0438	-0.0442	0.0069	0.0300	0.0003
			level 2 vs. level 3		
			α_0		
class 1	3.3360	2.2059	0.3966	1.4379	1.3682
class 2	0.2934	0.0936	-1.8840	1.4048	0.5097
class 3	1.9570	2.7255	0.7330	3.4921	2.9509
class 4	2.9822	1.0677	0.6211	4.6124	0.8161
class 5	-4.4138	-4.3121	-1.8665	-2.3500	-0.3425
class 6	0.4196	0.5214	-2.6498	-0.2605	0.8891
			α_{Lm}		
z_{1m}	-0.2110	-0.2738	-0.7938	-0.7829	-0.5860
z_{2m}	0.0179	-0.0076	0.0185	0.0075	-0.0115

Table 4: Values of β_0 and β_{pj} in 6 class case

	class 1 vs. class 6	class 2 vs. class 6	class 3 vs. class 6	class 4 vs. class 6	class 5 vs. class 6
			β_0		
	0.408	-0.313	-0.482	-0.033	-0.171
			β_{pj}		
x_{i1}	-0.464	-0.021	0.149	-0.187	0.158
x_{i2}	0.092	0.195	0.265	0.181	0.104

Table 5: Values of α_0 and α_{Lm} in 2 class case

	item 1	item 2	item 3	item 4	item 5
			level 1 vs. level 3		
			α_0		
class 1	3.0464	2.0611	-0.4656	2.8398	1.6822
class 2	-3.7435	-3.9859	-4.8386	-2.8929	-1.2178
			α_{Lm}		
z_{1m}	0.3906	0.3114	-0.0532	-0.4618	-0.2682
z_{2m}	0.0511	-0.0138	0.0243	0.0416	0.0111
			level 2 vs. level 3		
			α_0		
class 1	3.0959	1.6635	0.2100	2.9031	1.5960
class 2	-0.8657	-0.9226	-1.7795	-0.7267	0.3726
			α_{Lm}		
z_{1m}	0.5012	0.3124	-0.4417	0.1048	-0.2919
z_{2m}	0.0122	-0.0016	0.0192	0.0072	-0.0132

Table 6: Values of β_0 and β_{pj} in 2 class case

	class 1 vs. class 2
	β_0
	0.577
	β_{pj}
x_{i1}	0.307
x_{i2}	-0.103

Table 7: Average parameters estimations for 100 replication in 3-class model, N=100 (standard error in multinomial regression / sample standard error for 100 replication)

item 1

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-4.398**	**-3.374**	(1.870/2.397)	**-4.444**	(10.114/3.814)	**-1.046**	(1.414/1.599)	**-2.113**	(2.110/2.865)
class 1	**11.146**	**7.037**	(11.578/4.121)	**9.778**	(22.483/2.572)	**0.504**	(1.018/1.854)	**1.223**	(2.291/3.737)
class 2	**3.094**	**4.753**	(7.875/4.374)	**7.503**	(24.921/6.628)	**0.331**	(1.062/1.622)	**0.001**	(4.746/4.146)
sex	**0.394**	**0.036**	(1.221/1.485)	**0.864**	(4.422/1.785)	**-0.207**	(0.432/0.308)	**-0.200**	(0.434/0.314)
age	**0.075**	**0.049**	(0.041/0.055)	**0.051**	(0.233/0.061)	**0.035**	(0.024/0.016)	**0.034**	(0.025/0.016)

item 2

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-1.431**	**-1.653**	(3.715/2.790)	**-2.119**	(7.229/4.003)	**0.652**	(1.081/1.649)	**-0.270**	(3.098/2.657)
class 1	**4.897**	**5.647**	(3.660/3.383)	**7.651**	(8.370/5.032)	**0.517**	(1.190/1.675)	**0.959**	(3.636/3.535)
class 2	**3.046**	**3.590**	(4.185/3.723)	**2.839**	(15.678/7.492)	**0.129**	(1.008/1.657)	**-0.119**	(6.615/4.166)
sex	**-0.640**	**-0.869**	(1.098/1.251)	**-0.715**	(1.800/1.525)	**-0.798**	(0.449/0.399)	**-0.783**	(0.447/0.400)
age	**-0.053**	**-0.065**	(0.037/0.036)	**-0.088**	(0.046/0.049)	**-0.037**	(0.025/0.023)	**-0.038**	(0.026/0.022)

item 3

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-4.100**	**-5.315**	(4.196/3.307)	**-6.424**	(7.862/4.322)	**-1.496**	(1.265/1.631)	**-2.139**	(2.562/2.856)
class 1	**4.334**	**5.350**	(3.765/3.047)	**7.838**	(8.554/4.67)	**0.699**	(1.025/1.622)	**0.890**	(3.535/3.389)
class 2	**1.912**	**3.062**	(5.237/3.596)	**1.985**	(11.701/6.661)	**0.162**	(1.014/1.563)	**-0.466**	(4.488/4.115)
sex	**-0.058**	**0.006**	(0.586/0.589)	**-0.009**	(0.752/1.066)	**-0.423**	(0.454/0.344)	**-0.414**	(0.454/0.361)
age	**0.029**	**0.023**	(0.033/0.034)	**0.002**	(0.039/0.043)	**0.009**	(0.025/0.022)	**0.009**	(0.025/0.024)

item 4

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-20.669**	**-5.680**	(2.896/5.384)	**-7.280**	(7.075/5.89)	**-1.729**	(1.165/1.459)	**-2.758**	(1.148/2.796)
class 1	**13.044**	**5.924**	(3.184/3.306)	**7.740**	(10.621/4.95)	**0.722**	(0.943/1.597)	**1.320**	(2.356/3.597)
class 2	**12.970**	**4.081**	(3.548/3.901)	**7.454**	(22.45/6.907)	**1.494**	(0.782/1.615)	**-0.106**	(3.009/4.056)
sex	**1.086**	**-0.007**	(0.691/0.699)	**0.766**	(1.732/1.646)	**-0.274**	(0.424/0.160)	**-0.264**	(0.424/0.151)
age	**0.355**	**0.091**	(0.045/0.095)	**0.105**	(0.057/0.077)	**0.042**	(0.024/0.012)	**0.042**	(0.025/0.012)

item 5

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-1.104**	**-1.623**	(1.143/1.372)	**-0.825**	(1.937/1.996)	**-0.036**	(1.131/1.697)	**-0.096**	(2.354/2.985)

Table 7: Continued

class 1	**12.967**	**3.705**	(1.102/1.494)	**5.404**	(9.679/4.146)	**0.775**	(0.890/1.709)	**0.745**	(2.830/3.593)
class 2	**0.694**	**2.305**	(2.457/2.863)	**1.702**	(8.345/5.394)	**0.246**	(0.981/2.065)	**-0.777**	(4.312/4.546)
sex	**-0.856**	**-0.836**	(0.547/0.600)	**-0.710**	(0.576/0.609)	**-0.941**	(0.431/0.309)	**-0.937**	(0.432/0.315)
age	**0.019**	**0.024**	0.031/0.029)	**0.002**	(0.036/0.041)	**0.006**	(0.024/0.020)	**0.006**	(0.024/0.020)

Table 8: Average parameters estimations for 100 replication in 3-class model, N=100 (standard error in multinomial regression / sample standard error for 100 replication)

Class 1 vs. Class 3									
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.107**	**-0.499**	(0.544/0.483)	**-1.192**	(0.634/1.053)	**0.407**	(2.181/3.083)	**0.960**	(5.379/4.580)
occup	**-0.294**	**0.859**	(0.543/0.378)	**0.882**	(0.555/0.538)	**-0.068**	(2.070/3.031)	**-0.541**	(10.733/3.845)
dprime	**0.035**	**0.013**	(0.156/0.097)	**0.350**	(0.174/0.208)	**-0.010**	(0.241/0.297)	**-0.194**	(1.044/2.435)

Class 2 vs. Class 3									
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.716**	**-0.826**	(0.616/0.614)	**-1.220**	(0.710/1.091)	**0.083**	(2.126/0.532)	**0.027**	(8.513/4.745)
occup	**0.152**	**0.497**	(0.639/0.551)	**-0.075**	(6.056/2.408)	**-0.154**	(2.278/2.395)	**-0.564**	(16.384/5.529)
dprime	**-0.041**	**0.011**	(0.177/0.163)	**0.118**	(0.205/0.235)	**-0.011**	(0.248/0.272)	**-0.025**	(0.497/0.443)

Table 9: Average conditional Probability for 100 replication in 3-class model, N=100 (sample standard deviance in parentheses)

Class 1									
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
item 1									
Level 1	**1.000**	**0.953**	(0.042)	**0.972**	(0.031)	**0.637**	(0.099)	**0.614**	(0.193)
Level 2	**0.000**	**0.046**	(0.042)	**0.028**	(0.031)	**0.363**	(0.099)	**0.385**	(0.193)
item2									
Level 1	**0.802**	**0.735**	(0.092)	**0.775**	(0.115)	**0.395**	(0.092)	**0.380**	(0.128)
Level 2	**0.198**	**0.265**	(0.092)	**0.225**	(0.115)	**0.605**	(0.092)	**0.619**	(0.128)
item3									
Level 1	**0.760**	**0.675**	(0.108)	**0.702**	(0.143)	**0.347**	(0.089)	**0.326**	(0.119)
Level 2	**0.240**	**0.325**	(0.108)	**0.298**	(0.143)	**0.653**	(0.089)	**0.673**	(0.119)
item4									
Level 1	**0.990**	**0.926**	(0.049)	**0.920**	(0.058)	**0.571**	(0.093)	**0.544**	(0.173)
Level 2	**0.010**	**0.073**	(0.049)	**0.080**	(0.058)	**0.429**	(0.093)	**0.456**	(0.173)
item5									

Table 9: Continued									
Level 1	**1.000**	**0.880**	(0.071)	**0.905**	(0.076)	**0.595**	(0.086)	**0.583**	(0.162)
Level 2	**0.000**	**0.120**	(0.071)	**0.095**	(0.076)	**0.405**	(0.086)	**0.416**	(0.162)

				Class 2					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.797**	**0.780**	(0.184)	**0.810**	(0.285)	**0.611**	(0.139)	**0.494**	(0.278)
Level 2	**0.203**	**0.220**	(0.184)	**0.190**	(0.285)	**0.389**	(0.139)	**0.505**	(0.278)
				item2					
Level 1	**0.389**	**0.377**	(0.199)	**0.366**	(0.308)	**0.360**	(0.112)	**0.297**	(0.201)
Level 2	**0.611**	**0.623**	(0.199)	**0.634**	(0.308)	**0.640**	(0.112)	**0.702**	(0.201)
				item3					
Level 1	**0.220**	**0.285**	(0.168)	**0.268**	(0.286)	**0.292**	(0.105)	**0.235**	(0.163)
Level 2	**0.780**	**0.715**	(0.168)	**0.732**	(0.283)	**0.709**	(0.105)	**0.764**	(0.163)
				item4					
Level 1	**0.989**	**0.677**	(0.186)	**0.775**	(0.314)	**0.557**	(0.135)	**0.437**	(0.255)
Level 2	**0.011**	**0.323**	(0.186)	**0.225**	(0.314)	**0.443**	(0.135)	**0.562**	(0.255)
				item5					
Level 1	**0.442**	**0.609**	(0.194)	**0.547**	(0.311)	**0.520**	(0.131)	**0.430**	(0.249)
Level 2	**0.558**	**0.391**	(0.194)	**0.453**	(0.311)	**0.480**	(0.131)	**0.569**	(0.249)

				Class 3					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.151**	**0.201**	(0.099)	**0.191**	(0.114)	**0.536**	(0.144)	**0.441**	(0.231)
Level 2	**0.849**	**0.799**	(0.099)	**0.809**	(0.114)	**0.464**	(0.144)	**0.558**	(0.231)
				item2					
Level 1	**0.029**	**0.053**	(0.041)	**0.059**	(0.056)	**0.310**	(0.120)	**0.258**	(0.159)
Level 2	**0.971**	**0.947**	(0.041)	**0.941**	(0.056)	**0.690**	(0.120)	**0.741**	(0.159)
				item3					
Level 1	**0.040**	**0.041**	(0.036)	**0.039**	(0.044)	**0.255**	(0.090)	**0.228**	(0.165)
Level 2	**0.960**	**0.959**	(0.036)	**0.961**	(0.044)	**0.745**	(0.090)	**0.771**	(0.165)
				item4					
Level 1	**0.000**	**0.130**	(0.078)	**0.141**	(0.105)	**0.447**	(0.126)	**0.368**	(0.195)
Level 2	**1.000**	**0.870**	(0.078)	**0.859**	(0.105)	**0.553**	(0.126)	**0.632**	(0.195)
				item5					

Table 9: Continued									
Level 1	**0.284**	**0.243**	(0.085)	**0.287**	(0.130)	**0.483**	(0.138)	**0.476**	(0.235)
Level 2	**0.716**	**0.757**	(0.085)	**0.713**	(0.130)	**0.517**	(0.138)	**0.523**	(0.235)

Table 10: Average Latent Prevalence for 100 replication in 3-class model, N=100 (sample standard deviance in parentheses)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.300**	**0.346**	(0.074)	**0.366**	(0.131)	**0.346**	(0.142)	**0.419**	(0.203)
class2	**0.300**	**0.228**	(0.060)	**0.213**	(0.114)	**0.333**	(0.137)	**0.274**	(0.184)
class3	**0.400**	**0.426**	(0.066)	**0.421**	(0.088)	**0.320**	(0.128)	**0.307**	(0.179)

Table 11: Average Correlation Coefficients for 100 replication in3-class model,N=100 (total number of not NA values in parentheses)

	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.101**	(100)	**0.199**	(100)	**0.452**	(99)	**0.446**	(92)
class2	**0.088**	(100)	**0.303**	(100)	**0.407**	(96)	**0.402**	(76)
class3	**0.102**	(100)	**0.183**	(100)	**0.444**	(95)	**0.447**	(80)

Table 12: Average Match Proportions for 100 replication in 3-class model, N=100

	K_Corr	K_Cova	D_Corr	D_Cova
class1	**0.768**	**0.665**	**0.417**	**0.497**
class2	**0.398**	**0.445**	**0.397**	**0.324**
class3	**0.873**	**0.689**	**0.381**	**0.373**

Table 13: Average parameters estimations for 100 replication in 3-class model, N=500 (standard error in multinomial regression / sample standard error for 100 replication)

item 1

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-4.398**	**-3.495**	(0.683/1.207)	**-3.138**	(0.749/1.216)	**-1.130**	(0.435/0.311)	**-1.097**	(0.431/0.221)
class 1	**11.146**	**5.516**	(1.194/1.780)	**6.166**	(3.297/2.611)	**0.269**	(0.236/0.340)	**0.181**	(0.236/0.231)
class 2	**3.094**	**3.309**	(0.687/2.669)	**3.387**	(4.474/4.401)	**0.198**	(0.237/0.349)	**0.214**	(0.227/0.355)
sex	**0.394**	**0.428**	(0.293/0.402)	**0.421**	(0.321/0.357)	**0.196**	(0.187/0.163)	**0.199**	(0.187/0.154)
age	**0.075**	**0.052**	(0.017/0.022)	**0.045**	(0.02/0.02)	**0.039**	(0.012/0.011)	**0.039**	(0.012/0.011)

item 2

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-1.431**	**-0.445**	(0.618/0.824)	**-0.455**	(0.633/1.425)	**0.228**	(0.436/0.302)	**0.203**	(0.434/0.301)
class 1	**4.897**	**4.468**	(0.432/1.036)	**4.275**	(0.439/1.996)	**0.254**	(0.240/0.234)	**0.248**	(0.241/0.178)
class 2	**3.046**	**2.284**	(0.479/1.993)	**0.851**	(2.193/4.882)	**0.106**	(0.244/0.231)	**0.188**	(0.233/0.308)
sex	**-0.640**	**-0.625**	(0.266/0.324)	**-0.535**	(0.262/0.291)	**-0.347**	(0.190/0.127)	**-0.342**	(0.190/0.122)
age	**-0.053**	**-0.072**	(0.017/0.019)	**-0.068**	(0.017/0.022)	**-0.023**	(0.012/0.007)	**-0.023**	(0.012/0.007)

item 3

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-4.100**	**-3.639**	(0.680/0.739)	**-3.699**	(0.668/1.373)	**-1.631**	(0.458/0.396)	**-1.720**	(0.457/0.328)
class 1	**4.334**	**3.936**	(0.410/0.810)	**4.252**	(0.429/1.527)	**0.174**	(0.245/0.206)	**0.269**	(0.246/0.283)
class 2	**1.912**	**1.883**	(0.527/1.660)	**0.829**	(2.586/4.277)	**-0.058**	(0.251/0.304)	**0.134**	(0.240/0.361)
sex	**-0.058**	**0.002**	(0.259/0.248)	**0.021**	(0.255/0.339)	**0.022**	(0.195/0.149)	**0.028**	(0.196/0.144)
age	**0.029**	**0.013**	(0.016/0.015)	**0.013**	(0.016/0.018)	**0.024**	(0.012/0.010)	**0.024**	(0.012/0.010)

item 4

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-20.669**	**-4.473**	(0.759/1.456)	**-4.937**	(0.796/2.004)	**-1.455**	(0.434/0.235)	**-1.490**	(0.431/0.199)
class 1	**13.044**	**5.318**	(0.509/1.158)	**4.980**	(0.494/1.492)	**0.323**	(0.234/0.350)	**0.301**	(0.234/0.193)
class 2	**12.970**	**3.333**	(0.425/1.983)	**2.953**	(2.499/4.765)	**0.262**	(0.236/0.371)	**0.364**	(0.226/0.448)
sex	**1.086**	**0.321**	(0.302/0.283)	**0.407**	(0.305/0.38)	**0.158**	(0.185/0.062)	**0.166**	(0.186/0.066)
age	**0.355**	**0.066**	(0.018/0.022)	**0.082**	(0.02/0.027)	**0.044**	(0.012/0.005)	**0.044**	(0.012/0.005)

item 5

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-1.104**	**-0.977**	(0.508/0.884)	**-0.911**	(0.611/1.017)	**-0.026**	(0.423/0.415)	**-0.094**	(0.420/0.400)

Table 13: Continued

class 1	**12.967**	**3.493**	(0.330/1.499)	**4.260**	(1.110/2.730)	**0.186**	(0.231/0.318)	**0.344**	(0.233/0.293)
class 2	**0.694**	**2.088**	(0.761/3.576)	**1.898**	(0.820/3.560)	**0.084**	(0.233/0.300)	**0.186**	(0.222/0.402)
sex	**-0.856**	**-0.708**	(0.233/0.304)	**-0.727**	(0.246/0.274)	**-0.456**	(0.184/0.125)	**-0.455**	(0.184/0.114)
age	**0.019**	**0.007**	(0.013/0.012)	**0.006**	(0.015/0.016)	**0.011**	(0.011/0.012)	**0.011**	(0.011/0.011)

Table 14: Average parameters estimations for 100 replication in 3-class model, N=500 (standard error in multinomial regression / sample standard error for 100 replication)

		Class 1 vs. Class 3							
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.107**	**0.125**	(0.235/0.456)	**-1.001**	(0.26/0.639)	**0.076**	(0.255/0.435)	**-0.021**	(0.254/0.230)
occup	**-0.294**	**0.029**	(0.227/0.129)	**0.848**	(0.245/0.252)	**-0.137**	(0.257/0.420)	**-0.078**	(0.257/0.275)
dprime	**0.035**	**-0.098**	(0.067/0.043)	**0.298**	(0.074/0.09)	**0.001**	(0.074/0.083)	**-0.064**	(0.073/0.041)

		Class 2 vs. Class 3							
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.716**	**-0.693**	(0.299/0.587)	**-1.264**	(0.314/0.902)	**-0.072**	(0.263/0.516)	**-0.069**	(0.256/0.519)
occup	**0.152**	**0.093**	(0.285/0.240)	**0.346**	(0.347/0.459)	**-0.076**	(0.256/0.355)	**0.019**	(0.243/0.253)
dprime	**-0.041**	**-0.071**	(0.085/0.083)	**0.102**	(0.093/0.152)	**0.042**	(0.075/0.074)	**0.001**	(0.073/0.066)

Table 15: Average conditional Probability for 100 replication in 3-class model, N=500 (sample standard deviance in parentheses)

		Class 1							
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**1.000**	**0.968**	(0.029)	**0.977**	(0.016)	**0.628**	(0.042)	**0.615**	(0.029)
Level 2	**0.000**	**0.032**	(0.029)	**0.024**	(0.016)	**0.372**	(0.042)	**0.384**	(0.029)
				item2					
Level 1	**0.802**	**0.731**	(0.097)	**0.745**	(0.117)	**0.384**	(0.026)	**0.379**	(0.035)
Level 2	**0.198**	**0.269**	(0.097)	**0.255**	(0.117)	**0.616**	(0.026	**0.620**	(0.035)
				item3					
Level 1	**0.760**	**0.668**	(0.115)	**0.703**	(0.113)	**0.344**	(0.040)	**0.347**	(0.046)
Level 2	**0.240**	**0.332**	(0.115)	**0.297**	(0.113)	**0.656**	(0.040)	**0.652**	(0.046)
				item4					
Level 1	**0.990**	**0.949**	(0.039)	**0.935**	(0.025)	**0.596**	(0.046)	**0.586**	(0.043)
Level 2	**0.010**	**0.050**	(0.039)	**0.065**	(0.025)	**0.404**	(0.046)	**0.413**	(0.043)
				item5					

Table 15: Continued

Level 1	**1.000**	**0.881**	(0.236)	**0.903**	(0.066)	**0.571**	(0.036)	**0.594**	(0.043)
Level 2	**0.000**	**0.119**	(0.236)	**0.097**	(0.066)	**0.429**	(0.036)	**0.405**	(0.043)

				Class 2					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.797**	**0.747**	(0.227)	**0.724**	(0.310)	**0.611**	(0.047)	**0.619**	(0..043)
Level 2	**0.203**	**0.253**	(0.227)	**0.276**	(0.310)	**0.389**	(0.047)	**0.381**	(0.043)
				item2					
Level 1	**0.389**	**0.345**	(0.223)	**0.330**	(0.272)	**0.351**	(0.039)	**0.368**	(0.056)
Level 2	**0.611**	**0.655**	(0.223)	**0.670**	(0.272)	**0.649**	(0.039)	**0.632**	(0.056)
				item3					
Level 1	**0.220**	**0.259**	(0.235)	**0.248**	(0.248)	**0.286**	(0.043)	**0.316**	(0.046)
Level 2	**0.780**	**0.741**	(0.235)	**0.752**	(0.248)	**0.704**	(0.043)	**0.684**	(0.046)
				item4					
Level 1	**0.989**	**0.709**	(0.230)	**0.647**	(0.335)	**0.581**	(0.058)	**0.596**	(0.071)
Level 2	**0.011**	**0.291**	(0.230)	**0.353**	(0.335)	**0.419**	(0.058)	**0.404**	(0.071)
				item5					
Level 1	**0.442**	**0.573**	(0.236)	**0.567**	(0.271)	**0.549**	(0.053)	**0.555**	(0.056)
Level 2	**0.558**	**0.427**	(0.236)	**0.433**	(0.271)	**0.451**	(0.053)	**0.445**	(0.056)

				Class 3					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.151**	**0.206**	(0.094)	**0.230**	(0.111)	**0.565**	(0.050)	**0.573**	(0.045)
Level 2	**0.849**	**0.794**	(0.094)	**0.770**	(0.111)	**0.435**	(0.050)	**0.427**	(0.045)
				item2					
Level 1	**0.029**	**0.059**	(0.041)	**0.073**	(0.050)	**0.328**	(0.035)	**0.321**	(0.030)
Level 2	**0.971**	**0.941**	(0.041)	**0.927**	(0.050)	**0.672**	(0.035)	**0.679**	(0.030)
				Item3					
Level 1	**0.040**	**0.046**	(0.034)	**0.050**	(0.041)	**0.306**	(0.028)	**0.289**	(0.044)
Level 2	**0.960**	**0.954**	(0.034)	**0.950**	(0.041)	**0.694**	(0.0.28)	**0.711**	(0.044)
				item4					
Level 1	**0.000**	**0.147**	(0.093)	**0.177**	(0.104)	**0.519**	(0.049)	**0.515**	(0.037)
Level 2	**1.000**	**0.853**	(0.093)	**0.823**	(0.104)	**0.481**	(0.049)	**0.485**	(0.037)
				item5					

Table 15: Continued

Level 1	**0.284**	**0.275**	(0.085)	**0.282**	(0.083)	**0.529**	(0.052)	**0.507**	(0.053)
Level 2	**0.716**	**0.725**	(0.085)	**0.718**	(0.083)	**0.471**	(0.052)	**0.493**	(0.053)

Table 16: Average Latent Prevalences for 100 replication in 3-class model, N=500 (sample standard deviance in parentheses)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.300**	**0.372**	(0.081)	**0.453**	(0.114)	**0.336**	(0.085)	**0.286**	(0.054)
class2	**0.300**	**0.191**	(0.070)	**0.177**	(0.093)	**0.341**	(0.110)	**0.349**	(0.093)
class3	**0.400**	**0.436**	(0.078)	**0.370**	(0.084)	**0.323**	(0.079)	**0.365**	(0.074)

Table 17: Average Correlation Coefficients for 100 replication in 3-class model, N=500 (total number of not NA values in parentheses)

	K_Corr	K_Cova	D_Corr	D_Cova
class1	**0.049** (100)	**0.086** (100)	**0.446** (100)	**0.454** (100)
class2	**0.034** (100)	**0.091** (100)	**0.402** (100)	**0.414** (100)
class3	**0.067** (100)	**0.088** (100)	**0.459** (100)	**0.452** (100)

Table 18: Average Match Proportions for 100 replication in 3-class model, N=500

	K_Corr	K_Cova	D_Corr	D_Cova
class1	**0.832**	**0.753**	**0.390**	**0.336**
class2	**0.365**	**0.333**	**0.386**	**0.395**
class3	**0.889**	**0.556**	**0.365**	**0.411**

Table 19: Average parameters estimations for 100 replication in 6-class model, N=300 (standard error in multinomial regression / sample standard error for 100 replication)

item 1

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
level 1 vs. level 3									
intercept	**-6.029**	**0.228**	(1.143/4.641)	**1.844**	(1.25/9.46)	**-1.421**	(0.756/4.690)	**-1.950**	(0.720/6.775)
class 1	**11.924**	**10.125**	(0.849/7.939)	**11.540**	(0.512/13.163)	**2.035**	(0.717/4.504)	**-4.481**	(NaN/9.029)
class 2	**5.927**	**0.506**	(0.755/7.481)	**6.471**	(0.988/48.88)	**2.184**	(0.590/4.486)	**-0.252**	(0.413/9.207)
class 3	**7.983**	**2.315**	(1.035/7.127)	**3.102**	(0.711/11.635)	**-0.373**	(0.428/7.319)	**2.575**	(0.509/0.6856)
class 4	**6.883**	**4.363**	(0.744/8.950)	**3.248**	(NaN/12.919)	**2.199**	(0.591/5.253)	**-1.136**	(0.348/9.348)
class 5	**0.345**	**-4.520**	(0.917/5.341)	**-11.281**	(0.813/11.652)	**1.301**	(0.531/0.523)	**1.698**	(0.417/7.111)
sex	**-0.844**	**0.017**	(0.467/0.440)	**0.021**	(0.556/1.54)	**-0.254**	(0.309/0.194)	**-0.217**	(0.310/0.194)
age	**0.044**	**0.016**	(0.029/0.03)	**0.031**	(0.035/0.035)	**0.001**	(0.019/0.017)	**0.005**	(0.019/0.012)
level 2 vs. level 3									
intercept	**0.420**	**1.942**	(1.007/6.801)	**3.062**	(1.169/7.594)	**0.279**	(0.761/6.255)	**0.311**	(0.720/7.178)
class 1	**2.916**	**5.350**	(0.829/9.803)	**4.487**	(NaN/12.456)	**0.104**	(0.715/6.203)	**-3.378**	(NaN/13.058)
class 2	**-0.126**	**-1.058**	(0.671/9.132)	**4.154**	(0.678/48.578)	**0.214**	(0.587/6.106)	**-1.074**	(0.417/10.150)
class 3	**1.537**	**-0.150**	(0.732/11.962)	**1.403**	(0.666/11.047)	**1.246**	(0.434/9.852)	**0.138**	(0.501/7.324)
class 4	**2.563**	**3.095**	(0.680/10.402)	**4.037**	(0.768/12.696)	**0.226**	(0.587/6.238)	**1.376**	(0.334/10.629)
class 5	**-4.833**	**-4.877**	(0.615/7.053)	**-8.430**	(0.794/10.288)	**1.060**	(0.536/7.215)	**1.125**	(0.410/8.317)
sex	**-0.211**	**0.261**	(0.422/0.418	**0.321**	(0.523/1.471)	**0.100**	(0.313/0.172)	**0.050**	(0.315/0.236)
age	**0.018**	**0.0173**	(0.026/0.027)	**0.018**	(0.033/0.037)	**0.002**	(0.020/0.021)	**0.004**	(0.020/0.012)

item 2

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
level 1 vs. level 3									
intercept	**-2.221**	**-2.820**	(1.342/14.131)	**-6.534**	(1.081/14.855)	**-1.714**	(0.790/4.656)	**-2.027**	(0.757/5.136)
class 1	**6.873**	**9.238**	(1.734/15.122)	**14.547**	(0.786/15.798)	**0.746**	(0.685/7.840)	**-3.663**	(NaN/10.244)
class 2	**1.908**	**-0.024**	(1.128/18.510)	**4.786**	(0.634/19.049)	**2.447**	(0.598/4.988)	**0.682**	(0.451/8.902)
class 3	**5.070**	**4.445**	(0.909/13.492)	**9.973**	(0.639/15.834)	**-0.253**	(0.451/8.347)	**2.901**	(0.538/5.039)
class 4	**2.639**	**3.577**	(1.052/14.972)	**6.946**	(0.598/18.129)	**2.557**	(0.576/4.929)	**-0.886**	(0.360/8.260)
class 5	**-7.304**	**-4.673**	(0.746/15.287)	**-6.648**	(0.353/14.421)	**0.860**	(0.453/6.939)	**1.784**	(0.458/6.234)
sex	**-0.513**	**-0.165**	(0.471/0.483)	**-0.125**	(0.481/0.558)	**-0.297**	(0.327/0.288)	**-0.235**	(0.322/0.221)
age	**-0.044**	**-0.061**	(0.030/0.029)	**-0.043**	(0.03/0.031)	**-0.039**	(0.021/0.014)	**-0.041**	(0.021/0.016)
level 2 vs. level 3									
intercept	**0.521**	**1.054**	(0.867/6.678)	**0.688**	(0.859/9.632)	**-3.270**	(0.649/7.303)	**-1.001**	(0.641/7.636)
class 1	**1.685**	**-0.166**	(0.842/8.087)	**2.958**	(NaN/10.485)	**3.435**	(0.609/7.178)	**-5.546**	(NaN/14.505)
class 2	**-0.428**	**-1.859**	(0.564/9.838)	**-1.686**	(0.457/13.856)	**3.221**	(0.480/7.320)	**-2.845**	(0.367/11.516)

Table 19: Continued

class 3	**2.204**	**1.627**	(0.611/6.230)	**3.966**	(0.5/12.154)	**1.813**	(0.368/11.495)	**1.360**	(0.439/7.413)
class 4	**0.546**	**-0.389**	(0.579/8.945)	**0.820**	(0.511/13.506)	**3.352**	(0.474/7.451)	**-0.126**	(0.297/13.883)
class 5	**-4.833**	**-3.912**	(0.531/7.856)	**-5.827**	(0.507/11.646)	**4.480**	(0.470/7.964)	**1.222**	(0.376/9.214)
sex	**-0.274**	**-0.086**	(0.327/0.327)	**-0.056**	(0.372/0.448)	**-0.164**	(0.267/0.328)	**-0.084**	(0.268/0.272)
age	**-0.008**	**-0.013**	(0.020/0.021)	**-0.008**	(0.023/0.026)	**0.001**	(0.017/0.008)	**-0.006**	(0.017/0.017)

					item 3				
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
					level 1 vs. level 3				
intercept	**-7.445**	**-5.547**	(1.160/11.578)	**-13.160**	(0.847/14.946)	**-3.606**	(0.837/4.648)	**-4.066**	(0.790/5.573)
class 1	**8.520**	**7.772**	(0.689/12.501)	**14.795**	(0.548/14.646)	**2.238**	(0.775/4.582)	**-3.194**	(NaN/8.797)
class 2	**2.959**	**0.769**	(0.666/15.504)	**5.581**	(0.344/18.171)	**2.020**	(0.643/4.452)	**-0.994**	(0.433/9.286)
class 3	**4.838**	**2.886**	(0.760/11.698)	**8.864**	(0.461/15.489)	**-1.453**	(0.505/8.607)	**2.840**	(0.594/5.490)
class 4	**7.931**	**7.152**	(0.996/13.821)	**9.811**	(0.429/16.866)	**2.355**	(0.609/4.456)	**-1.398**	(0.358/8.872)
class 5	**3.620**	**-3.671**	(0.672/14.895)	**1.398**	(0.336/13.851)	**1.206**	(0.581/6.398)	**1.582**	(0.495/7.173)
sex	**-0.450**	**-0.408**	(0.476/0.497)	**-0.047**	(0.442/0.553)	**-0.339**	(0.342/0.344)	**-0.456**	(0.342/0.263)
age	**0.007**	**-0.008**	(0.030/0.033)	**0.003**	(0.028/0.034)	**0.010**	(0.021/0.016)	**0.002**	(0.022/0.018)
					level 2 vs. level 3				
intercept	**-2.650**	**-2.022**	(0.795/4.428)	**-5.241**	(0.793/12.547)	**-3.155**	(0.652/6.175)	**-4.931**	(0.659/7.284)
class 1	**3.046**	**2.183**	(0.525/5.008)	**5.938**	(0.638/12.714)	**2.642**	(0.624/5.736)	**-0.473**	(0.385/17.349)
class 2	**0.766**	**-0.196**	(0.528/7.456)	**2.768**	(0.533/10.714)	**2.395**	(0.508/5.800)	**0.314**	(0.379/11.427)
class 3	**3.383**	**3.741**	(0.514/10.244)	**6.147**	(0.542/13.399)	**-1.952**	(0.367/10.719)	**4.195**	(0.458/7.407)
class 4	**3.271**	**1.821**	(0.758/6.941)	**4.325**	(0.554/13.633)	**2.864**	(0.470/5.689)	**4.088**	(0.314/18.942)
class 5	**0.783**	**-0.725**	(0.611/5.254)	**2.992**	(0.616/12.729)	**1.273**	(0.505/7.865)	**2.393**	(0.403/9.624)
sex	**-0.794**	**-0.745**	(0.311/0.321)	**-0.547**	(0.319/0.378)	**-0.541**	(0.271/0.207)	**-0.527**	(0.272/0.279)
age	**0.019**	**0.008**	(0.019/0.016)	**0.014**	(0.020/0.020)	**0.006**	(0.017/0.018)	**0.017**	(0.017/0.011)

					item 4				
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
					level 1 vs. level 3				
intercept	**-6.883**	**-0.593**	(1.073/7.724)	**-0.232**	(1.11/8.601)	**-2.168**	(0.735/5.196)	**-4.291**	(0.715/8.172)
class 1	**10.946**	**7.584**	(0.837/11.171)	**14.228**	(NaN/23.778)	**2.835**	(0.720/5.599)	**-4.582**	(NaN/11.582)
class 2	**7.782**	**2.785**	(0.699/8.628)	**5.210**	(0.597/12.414)	**2.730**	(0.564/5.010)	**2.848**	(0.419/11.057)
class 3	**9.085**	**3.679**	(0.755/9.217)	**5.279**	(0.78/11.649)	**-0.979**	(0.422/9.450)	**4.661**	(0.504/8.388)
class 4	**11.649**	**6.405**	(0.669/9.924)	**6.106**	(0.577/11.194)	**2.721**	(0.560/4.983)	**-1.083**	(0.344/12.796)
class 5	**2.957**	**-2.751**	(0.806/9.457)	**-6.178**	(0.687/9.665)	**0.859**	(0.510/7.327)	**3.433**	(0.424/9.563)
sex	**-1.833**	**-1.065**	(0.446/0.436)	**-1.053**	(0.516/0.749	**-0.740**	(0.304/0.218)	**-0.778**	(0.306/0.267)

Table 19: Continued

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
age	**0.030**	**-0.002**	(0.027/0.026)	**0.020**	(0.031/0.031)	**0.011**	(0.019/0.019)	**0.015**	(0.019/0.012)
				level 2 vs. level 3					
intercept	**-0.261**	**-0.134**	(0.983/7.815)	**0.123**	(1.025/6.465)	**-2.978**	(0.742/7.243)	**-2.885**	(0.736/5.549)
class 1	**1.698**	**4.229**	(0.868/9.536)	**4.434**	(NaN/21.747)	**3.343**	(0.740/7.212)	**2.749**	(NaN/17.709)
class 2	**1.665**	**2.729**	(0.642/8.727)	**3.804**	(0.547/10.582)	**3.385**	(0.560/6.712)	**1.473**	(0.418/9.460)
class 3	**3.753**	**3.838**	(0.702/9.608)	**5.606**	(0.747/10.134)	**1.914**	(0.414/10.705)	**3.414**	(0.517/5.733)
class 4	**4.873**	**5.845**	(0.660/11.444)	**5.971**	(0.546/10.546)	**3.442**	(0.553/6.584)	**-1.138**	(0.366/10.816)
class 5	**-2.089**	**-1.946**	(0.601/9.057)	**-4.647**	(0.606/10.304)	**1.517**	(0.513/8.706)	**3.192**	(0.436/7.129)
sex	**-0.783**	**-0.266**	(0.405/0.398)	**-0.289**	(0.479/0.632)	**0.178**	(0.313/0.280)	**0.076**	(0.312/0.169)
age	**0.008**	**-0.007**	(0.025/0.024)	**0.001**	(0.029/0.027)	**-0.002**	(0.020/0.019)	**-0.005**	(0.019/0.013)

				item 5					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				level 1 vs. level 3					
intercept	**0.199**	**0.189**	(0.914/5.311)	**2.523**	(0.939/10.278)	**-1.282**	(0.722/5.814)	**-0.575**	(0.703/6.285)
class 1	**3.300**	**3.527**	(0.764/5.816)	**4.531**	(0.908/12.099)	**2.853**	(0.667/5.790)	**0.030**	(NaN/11.977)
class 2	**0.106**	**-0.942**	(0.557/8.913)	**0.139**	(0.594/14.111)	**2.789**	(0.553/5.794)	**-0.423**	(0.398/9.529)
class 3	**1.796**	**1.999**	(0.622/5.739)	**1.275**	(0.616/11.492)	**1.824**	(0.400/9.299)	**1.633**	(0.491/6.409)
class 4	**0.370**	**0.407**	(0.643/8.508)	**-0.008**	(0.515/12.751)	**2.601**	(0.521/5.602)	**-0.141**	(0.332/7.596)
class 5	**-2.095**	**-1.770**	(0.662/5.777)	**-5.033**	(0.647/11.443)	**4.013**	(0.526/6.219)	**0.016**	(0.413/8.133)
sex	**-0.767**	**-0.531**	(0.366/0.368)	**-0.570**	(0.385/0.427)	**-0.662**	(0.296/0.234)	**-0.570**	(0.296/0.227)
age	**0.000**	**-0.008**	(0.023/0.024)	**0.003**	(0.024/0.026)	**-0.022**	(0.018/0.011)	**-0.010**	(0.018/0.016)
				level 2 vs. level 3					
intercept	**0.889**	**0.665**	(0.884/6.007)	**3.891**	(0.917/11.262)	**-1.697**	(0.762/6.409)	**1.916**	(0.738/5.565)
class 1	**0.479**	**0.843**	(0.842/6.621)	**-0.487**	(0.928/10.74)	**2.816**	(0.730/6.400)	**-4.647**	(NaN/10.568)
class 2	**-0.379**	**-1.431**	(0.556/8.146)	**-1.726**	(0.58/13.667)	**3.281**	(0.561/6.451)	**-2.552**	(0.420/8.440)
class 3	**20.620**	**3.065**	(0.602/7.739)	**0.590**	(0.608/11.761)	**2.338**	(0.426/9.221)	**-0.882**	(0.515/5.637)
class 4	**-0.073**	**-0.282**	(0.654/7.729)	**-0.948**	(0.582/13.179)	**2.932**	(0.542/6.116)	**3.645**	(0.345/15.829)
class 5	**-1.232**	**-1.096**	(0.556/6.107)	**-4.696**	(0.593/11.888)	**1.183**	(0.441/8.362)	**-2.472**	(0.433/7.292)
sex	**-0.586**	**-0.511**	(0.352/0.423)	**-0.494**	(0.371/0.473)	**-0.482**	(0.312/0.216)	**-0.392**	(0.312/0.270)
age	**-0.011**	**-0.011**	(0.022/0.025)	**-0.012**	(0.023/0.024)	**-0.027**	(0.020/0.013)	**-0.019**	(0.019/0.017)

Table 20: Average parameters estimations for 100 replication in 6-class model, N=300 (standard error in multinomial regression / sample standard error for 100 replication)

				Class 1 vs. Class 6					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**0.408**	**0.316**	(0.419/0.700)	**-0.484**	(0.523/1.157)	**0.915**	(1.568/3.492)	**-1.598**	(6.303/8.167)

Table 20: Continued

occup	**-0.464**	**0.005**	(0.493/0.429)	**0.666**	(1.754/1.231)	**1.623**	(2.516/2.942)	**1.732**	(16.376/6.278)
dprime	**0.092**	**0.031**	(0.127/0.132)	**0.407**	(0.157/0.233)	**-0.324**	(0.407/0.563)	**0.021**	(0.729/1.164)

Class 2 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.313**	**0.010**	(0.453/0.819)	**0.023**	(0.518/1.505)	**1.446**	(1.524/3.674)	**-0.099**	(3.079/7.510)
occup	**-0.021**	**-0.026**	(0.545/0.416)	**-0.081**	(2.769/1.725)	**1.728**	(2.423/3.366)	**-0.755**	(33.740/5.402)
dprime	**0.195**	**-0.003**	(0.139/0.134)	**0.064**	(0.168/0.331)	**-0.217**	(0.386/0.559)	**-0.007**	(0.783/1.385)

Class 3 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.482**	**0.361**	(0.417/0.744)	**0.180**	(0.501/1.492)	**0.080**	(2.099/4.759)	**2.107**	(1.797/4.979)
occup	**0.149**	**0.054**	(0.504/0.496)	**-0.007**	(1.799/1.303)	**-0.241**	(14.844/4.921)	**1.341**	(12.134/2.990)
dprime	**0.265**	**-0.040**	(0.129/0.125)	**0.190**	(0.157/0.277)	**-0.260**	(0.619/0.717)	**-0.210**	(0.470/1.010)

Class 4 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.033**	**0.021**	(0.453/0.780)	**-0.066**	(0.524/1.271)	**1.535**	(1.514/3.573)	**-11.268**	(15.798/48.109)
occup	**-0.187**	**0.004**	(0.548/0.495)	**0.186**	(1.838/1.209)	**1.621**	(2.414/3.221)	**3.888**	(27.479/22.352)
dprime	**0.181**	**-0.028**	(0.140/0.137)	**0.128**	(0.168/0.27)	**-0.232**	(0.385/0.566)	**1.088**	(5.048/5.683)

Class 5 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.171**	**0.217**	(0.421/0.684)	**0.855**	(0.43/1.098)	**-15.545**	(18.816/53.500)	**-10.414**	(1.951/55.382)
occup	**0.158**	**0.064**	(0.484/0.417)	**0.062**	(1.781/1.25)	**7.307**	(19.687/18.638)	**7.271**	(13.701/26.852)
dprime	**0.104**	**0.074**	(0.127/0.108)	**-0.092**	(0.146/0.221)	**2.049**	(7.504/7.235)	**1.149**	(0.516/5.825)

Table 21: Average conditional Probability for 100 replication in 6-class model, N=300 (sample standard deviance in parentheses)

Class 1

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.955**	**0.837**	(0.086)	**0.914**	(0.077)	**0.386**	(0.117)	**0.314**	(0.214)
Level 2	**0.044**	**0.152**	(0.082)	**0.079**	(0.068)	**0.377**	(0.131)	**0.323**	(0.270)
Level 3	**0.001**	**0.012**	(0.019)	**0.007**	(0.014)	**0.236**	(0.074)	**0.361**	(0.338)
				item 2					

Table 21: Continued

Level 1	**0.727**	**0.693**	(0.177)	**0.758**	(0.136)	**0.195**	(0.126)	**0.173**	(0.125)
Level 2	**0.235**	**0.241**	(0.144)	**0.200**	(0.111)	**0.418**	(0.110)	**0.322**	(0.241)
Level 3	**0.038**	**0.067**	(0.055)	**0.042**	(0.043)	**0.386**	(0.132)	**0.503**	(0.311)
					item 3				
Level 1	**0.508**	**0.445**	(0.222)	**0.531**	(0.155)	**0.163**	(0.047)	**0.131**	(0.106)
Level 2	**0.318**	**0.362**	(0.167)	**0.327**	(0.132)	**0.293**	(0.074)	**0.265**	(0.235)
Level 3	**0.174**	**0.194**	(0.100)	**0.142**	(0.072)	**0.542**	(0.095)	**0.603**	(0.274)
					item 4				
Level 1	**0.929**	**0.830**	(0.086)	**0.900**	(0.083)	**0.420**	(0.133)	**0.283**	(0.204)
Level 2	**0.056**	**0.147**	(0.081)	**0.087**	(0.076)	**0.343**	(0.151)	**0.334**	(0.268)
Level 3	**0.015**	**0.024**	(0.019)	**0.013**	(0.019)	**0.235**	(0.105)	**0.382**	(0.331)
					item 5				
Level 1	**0.883**	**0.794**	(0.093)	**0.861**	(0.088)	**0.445**	(0.105)	**0.397**	(0.275)
Level 2	**0.078**	**0.143**	(0.077)	**0.095**	(0.070)	**0.266**	(0.053)	**0.239**	(0.174)
Level 3	**0.039**	**0.064**	(0.042)	**0.044**	(0.036)	**0.287**	(0.085)	**0.362**	(0.340)

					Class 2				
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
					item 1				
Level 1	**0.439**	**0.355**	(0.182)	**0.457**	(0.273)	**0.398**	(0.025)	**0.302**	(0.190)
Level 2	**0.384**	**0.445**	(0.182)	**0.378**	(0.226)	**0.386**	(0.046)	**0.373**	(0.266)
Level 3	**0.177**	**0.201**	(0.207)	**0.165**	(0.260)	**0.214**	(0.049)	**0.324**	(0.304)
					item 2				
Level 1	**0.071**	**0.102**	(0.114)	**0.211**	(0.282)	**0.205**	(0.042)	**0.203**	(0.213)
Level 2	**0.397**	**0.423**	(0.196)	**0.363**	(0.245)	**0.380**	(0.043)	**0.295**	(0.188)
Level 3	**0.532**	**0.475**	(0.230)	**0.427**	(0.338)	**0.414**	(0.070)	**0.501**	(0.277)
					item 3				
Level 1	**0.009**	**0.102**	(0.175)	**0.104**	(0.196)	**0.149**	(0.060)	**0.093**	(0.073)
Level 2	**0.156**	**0.246**	(0.153)	**0.277**	(0.206)	**0.263**	(0.079)	**0.239**	(0.172)
Level 3	**0.834**	**0.653**	(0.207)	**0.619**	(0.266)	**0.586**	(0.045)	**0.667**	(0.211)
					item 4				
Level 1	**0.361**	**0.345**	(0.363)	**0.409**	(0.252)	**0.400**	(0.049)	**0.316**	(0.223)
Level 2	**0.498**	**0.467**	(0.380)	**0.443**	(0.232)	**0.364**	(0.055)	**0.331**	(0.225)
Level 3	**0.141**	**0.188**	(0.288)	**0.148**	(0.200)	**0.235**	(0.049)	**0.352**	(0.295)
					item 5				
Level 1	**0.335**	**0.348**	(0.196)	**0.411**	(0.280)	**0.368**	(0.058)	**0.344**	(0.228)
Level 2	**0.306**	**0.288**	(0.163)	**0.302**	(0.196)	**0.376**	(0.085)	**0.285**	(0.221)

Table 21: Continued

Level 3	**0.359**	**0.364**	(0.273)	**0.287**	(0.276)	**0.2549**	(0.054)	**0.370**	(0.289)

	Class 3								
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.609**	**0.431**	(0.185)	**0.505**	(0.192)	**0.304**	(0.217)	**0.406**	(0.063)
Level 2	**0.360**	**0.471**	(0.176)	**0.430**	(0.181)	**0.441**	(0.315)	**0.381**	(0.060)
Level 3	**0.031**	**0.098**	(0.175)	**0.065**	(0.145)	**0.254**	(0.280)	**0.212**	(0.059)
				item 2					
Level 1	**0.217**	**0.152**	(0.127)	**0.194**	(0.175)	**0.157**	(0.118)	**0.216**	(0.050)
Level 2	**0.714**	**0.627**	(0.160)	**0.653**	(0.206)	**0.397**	(0.279)	**0.411**	(0.050)
Level 3	**0.069**	**0.222**	(0.171)	**0.153**	(0.176)	**0.444**	(0.320)	**0.372**	(0.051)
				item 3					
Level 1	**0.020**	**0.093**	(0.112)	**0.088**	(0.140)	**0.109**	(0.089)	**0.137**	(0.046)
Level 2	**0.705**	**0.480**	(0.177)	**0.573**	(0.206)	**0.225**	(0.176)	**0.333**	(0.064)
Level 3	**0.275**	**0.427**	(0.188)	**0.340**	(0.202)	**0.665**	(0.245)	**0.528**	(0.083)
				item 4					
Level 1	**0.242**	**0.356**	(0.172)	**0.346**	(0.210)	**0.272**	(0.197)	**0.384**	(0.064)
Level 2	**0.732**	**0.541**	(0.174)	**0.589**	(0.217)	**0.360**	(0.282)	**0.372**	(0.076)
Level 3	**0.026**	**0.104**	(0.136)	**0.065**	(0.123)	**0.366**	(0.344)	**0.242**	(0.037)
				item 5					
Level 1	**0.319**	**0.359**	(0.174)	**0.368**	(0.183)	**0.378**	(0.275)	**0.399**	(0.046)
Level 2	**0.618**	**0.510**	(0.189)	**0.503**	(0.194)	**0.330**	(0.282)	**0.325**	(0.062)
Level 3	**0.063**	**0.133**	(0.111)	**0.129**	(0.138)	**0.291**	(0.280)	**0.274**	(0.067)

	Class 4								
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.164**	**0.419**	(0.210)	**0.408**	(0.271)	**0.404**	(0.074)	**0.235**	(0.181)
Level 2	**0.811**	**0.498**	(0.201)	**0.493**	(0.273)	**0.383**	(0.094)	**0.474**	(0.312)
Level 3	**0.025**	**0.083**	(0.162)	**0.099**	(0.230)	**0.212**	(0.062)	**0.290**	(0.322)
				item 2					
Level 1	**0.085**	**0.192**	(0.188)	**0.202**	(0.227)	**0.223**	(0.046)	**0.143**	(0.114)
Level 2	**0.607**	**0.525**	(0.200)	**0.520**	(0.251)	**0.397**	(0.078)	**0.405**	(0.305)
Level 3	**0.307**	**0.283**	(0.191)	**0.278**	(0.286)	**0.379**	(0.080)	**0.451**	(0.317)
				item 3					

Table 21: Continued

Level 1	**0.330**	**0.311**	(0.356)	**0.204**	(0.252)	**0.167**	(0.045)	**0.124**	(0.101)
Level 2	**0.466**	**0.337**	(0.241)	**0.397**	(0.235)	**0.339**	(0.063)	**0.293**	(0.282)
Level 3	**0.203**	**0.353**	(0.254)	**0.399**	(0.277)	**0.492**	(0.047)	**0.582**	(0.298)
					item 4				
Level 1	**0.581**	**0.487**	(0.234)	**0.440**	(0.246)	**0.393**	(0.079)	**0.274**	(0.219)
Level 2	**0.414**	**0.412**	(0.215)	**0.419**	(0.219)	**0.376**	(0.093)	**0.275**	(0.247)
Level 3	**0.005**	**0.101**	(0.142)	**0.141**	(0.256)	**0.229**	(0.058)	**0.449**	(0.375)
					item 5				
Level 1	**0.360**	**0.422**	(0.226)	**0.404**	(0.241)	**0.380**	(0.051)	**0.353**	(0.288)
Level 2	**0.343**	**0.300**	(0.163)	**0.341**	(0.217)	**0.325**	(0.046)	**0.457**	(0.342)
Level 3	**0.296**	**0.279**	(0.233)	**0.255**	(0.264)	**0.293**	(0.056)	**0.189**	(0.154)

					Class 5				
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
					item 1				
Level 1	**0.009**	**0.045**	(0.034)	**0.025**	(0.029)	**0.345**	(0.128)	**0.344**	(0.131)
Level 2	**0.019**	**0.189**	(0.102)	**0.098**	(0.070)	**0.404**	(0.219)	**0.412**	(0.208)
Level 3	**0.972**	**0.767**	(0.119)	**0.877**	(0.086)	**0.250**	(0.114)	**0.242**	(0.090)
					item 2				
Level 1	**0.000**	**0.010**	(0.010)	**0.003**	(0.008)	**0.176**	(0.100)	**0.177**	(0.075)
Level 2	**0.009**	**0.120**	(0.077)	**0.075**	(0.057)	**0.445**	(0.233)	**0.388**	(0.171)
Level 3	**0.991**	**0.870**	(0.079)	**0.922**	(0.060)	**0.377**	(0.146)	**0.434**	(0.169)
					item 3				
Level 1	**0.018**	**0.013**	(0.014)	**0.014**	(0.019)	**0.164**	(0.077)	**0.131**	(0.067)
Level 2	**0.157**	**0.114**	(0.056)	**0.141**	(0.065)	**0.287**	(0.163)	**0.272**	(0.115)
Level 3	**0.825**	**0.873**	(0.056)	**0.845**	(0.070)	**0.548**	(0.190)	**0.596**	(0.152)
					item 4				
Level 1	**0.019**	**0.054**	(0.038)	**0.036**	(0.036)	**0.325**	(0.124)	**0.326**	(0.120)
Level 2	**0.075**	**0.156**	(0.086)	**0.095**	(0.059)	**0.275**	(0.111)	**0.368**	(0.176)
Level 3	**0.907**	**0.790**	(0.100)	**0.869**	(0.076)	**0.399**	(0.215)	**0.305**	(0.181)
					item 5				
Level 1	**0.070**	**0.108**	(0.053)	**0.104**	(0.053)	**0.482**	(0.243)	**0.343**	(0.125)
Level 2	**0.248**	**0.260**	(0.097)	**0.235**	(0.098)	**0.271**	(0.164)	**0.298**	(0.109)
Level 3	**0.682**	**0.632**	(0.118)	**0.661**	(0.123)	**0.246**	(0.108)	**0.358**	(0.224)

Table 21: Continued

				Class 6					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.002**	**0.249**	(0.215)	**0.282**	(0.292)	**0.286**	(0.157)	**0.264**	(0.168)
Level 2	**0.709**	**0.522**	(0.200)	**0.494**	(0.288)	**0.434**	(0.247)	**0.420**	(0.259)
Level 3	**0.289**	**0.230**	(0.201)	**0.224**	(0.270)	**0.279**	(0.267)	**0.315**	(0.311)
				item 2					
Level 1	**0.009**	**0.114**	(0.192)	**0.121**	(0.246)	**0.142**	(0.083)	**0.150**	(0.097)
Level 2	**0.529**	**0.430**	(0.199)	**0.438**	(0.257)	**0.345**	(0.191)	**0.405**	(0.260)
Level 3	**0.462**	**0.456**	(0.244)	**0.441**	(0.286)	**0.512**	(0.260)	**0.444**	(0.277)
				item 3					
Level 1	**0.001**	**0.120**	(0.252)	**0.065**	(0.207)	**0.114**	(0.067)	**0.127**	(0.090)
Level 2	**0.080**	**0.218**	(0.153)	**0.182**	(0.188)	**0.254**	(0.142)	**0.197**	(0.132)
Level 3	**0.919**	**0.663**	(0.252)	**0.753**	(0.259)	**0.631**	(0.201)	**0.675**	(0.199)
				item 4					
Level 1	**0.001**	**0.252**	(0.214)	**0.257**	(0.285)	**0.281**	(0.154)	**0.289**	(0.181)
Level 2	**0.400**	**0.390**	(0.207)	**0.341**	(0.238)	**0.297**	(0.160)	**0.252**	(0.170)
Level 3	**0.600**	**0.359**	(0.268)	**0.401**	(0.331)	**0.420**	(0.307)	**0.458**	(0.329)
				item 5					
Level 1	**0.272**	**0.305**	(0.210)	**0.330**	(0.238)	**0.321**	(0.183)	**0.348**	(0.227)
Level 2	**0.404**	**0.378**	(0.215)	**0.396**	(0.233)	**0.245**	(0.146)	**0.392**	(0.286)
Level 3	**0.324**	**0.318**	(0.252)	**0.274**	(0.245)	**0.432**	(0.302)	**0.259**	(0.217)

Table 22: Average Latent Prevalences for 100 replication in 6-class model, N=300 (sample standard deviance in parentheses)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.200**	**0.204**	(0.065)	**0.217**	(0.086)	**0.102**	(0.092)	**0.124**	(0.130)
class2	**0.200**	**0.138**	(0.050)	**0.132**	(0.079)	**0.216**	(0.150)	**0.116**	(0.110)
class3	**0.150**	**0.179**	(0.054)	**0.199**	(0.097)	**0.139**	(0.186)	**0.234**	(0.144)
class4	**0.150**	**0.135**	(0.054)	**0.156**	(0.101)	**0.206**	(0.127)	**0.123**	(0.133)
class5	**0.150**	**0.204**	(0.049)	**0.183**	(0.051)	**0.168**	(0.170)	**0.251**	(0.168)
class6	**0.150**	**0.142**	(0.054)	**0.113**	(0.060)	**0.169**	(0.159)	**0.152**	(0.142)

Table 23: Average Correlation Coefficients for 100 replication in 6-class model, N=300 (total number of not NA values in parentheses)

	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.133**	(100)	**0.168**	(100)	**0.316**	(100)	**0.308**	(80)
class2	**0.136**	(100)	**0.198**	(100)	**0.286**	(100)	**0.316**	(80)
class3	**0.112**	(100)	**0.149**	(100)	**0.312**	(96)	**0.296**	(98)
class4	**0.136**	(100)	**0.184**	(100)	**0.301**	(100)	**0.267**	(88)
class5	**0.185**	(100)	**0.226**	(100)	**0.358**	(96)	**0.317**	(90)
class6	**0.157**	(100)	**0.216**	(100)	**0.281**	(98)	**0.283**	(90)

Table 24: Average Match Proportions for 100 replication in 6-class model, N=300

	K_Corr	K_Cova	D_Corr	D_Cova
class1	**0.651**	**0.636**	**0.122**	**0.150**
class2	**0.253**	**0.244**	**0.280**	**0.149**
class3	**0.396**	**0.424**	**0.173**	**0.303**
class4	**0.270**	**0.267**	**0.273**	**0.172**
class5	**0.782**	**0.798**	**0.218**	**0.306**
class6	**0.284**	**0.294**	**0.233**	**0.218**

Table 25: Average parameters estimations for 100 replication in 6-class model, N=1000
(standard error in multinomial regression / sample standard error for 100 replication)

item 1									
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
level 1 vs. level 3									
intercept	-6.029	-1.350	(0.712/3.492)	-0.415	(0.753/4.672)	-0.973	(0.758/3.042)	-2.617	(0.754/5.633)
class 1	11.924	5.659	(0.828/4.682)	9.291	(0.603/7.259)	0.589	(0.464/4.737)	2.519	(0.508/6.184)
class 2	5.927	1.459	(0.547/5.186)	1.597	(0.501/7.569)	1.648	(0.658/2.905)	3.364	(0.647/5.733)
class 3	7.983	3.223	(0.571/3.695)	2.845	(0.525/11.486)	-1.600	(0.463/8.725)	2.667	(0.453/6.562)
class 4	6.883	2.915	(0.620/6.534)	2.536	(0.511/8.388)	-1.838	(0.413/7.876)	3.135	(0.629/5.459)
class 5	0.345	-3.843	(0.599/4.203)	-6.763	(NaN/7.542)	1.254	(0.453/2.468)	2.189	(0.459/6.131)
sex	-0.844	0.041	(0.277/0.401)	-0.017	(0.297/0.476)	-0.287	(0.312/0.144)	-0.238	(0.307/0.141)
age	0.044	0.035	(0.018/0.020)	0.041	(0.019/0.023)	0.002	(0.019/0.016)	0.001	(0.019/0.012)
level 2 vs. level 3									
intercept	0.420	-0.098	(0.637/3.172)	0.863	(0.666/4.140)	-3.083	(0.741/5.108)	0.690	(0.765/7.613)
class 1	2.916	2.536	(0.833/4.525)	5.074	(0.476/6.726)	0.102	(0.454/8.535)	1.867	(0.534/9.267)
class 2	-0.126	0.594	(0.494/4.979)	0.332	(0.409/6.990)	-2.301	(0.633/5.383)	-0.179	(0.671/7.600)
class 3	1.537	2.386	(0.533/3.561)	1.924	(0.484/11.036)	-5.546	(0.448/7.925)	2.049	(0.458/9.133)
class 4	2.563	3.101	(0.591/6.670)	2.093	(0.471/7.575)	-2.284	(0.417/10.453)	-0.128	(0.629/7.504)
class 5	-4.833	-3.281	(0.426/4.406)	-4.573	(0.500/4.285)	-2.498	(0.418/5.423)	1.835	(0.454/9.501)
sex	-0.211	0.242	(0.254/0.392)	0.202	(0.276/0.430)	0.035	(0.315/0.231)	0.027	(0.314/0.220)
age	0.018	0.018	(0.016/0.018)	0.025	(0.018/0.021)	-0.007	(0.019/0.012)	-0.003	(0.019/0.011)

item 2									
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
level 1 vs. level 3									
intercept	-2.221	-2.008	(0.771/5.309)	-2.634	(0.672/5.215)	-1.663	(0.781/4.231)	-2.116	(0.740/5.209)
class 1	6.873	7.532	(0.687/7.625)	8.602	(0.543/7.344)	-0.207	(0.445/8.705)	2.339	(0.454/5.541)
class 2	1.908	0.827	(0.669/7.332)	0.766	(0.498/8.021)	2.108	(0.654/4.801)	2.075	(0.762/5.240)
class 3	5.070	3.128	(0.646/5.314)	3.895	(0.496/6.749)	1.210	(0.471/5.195)	0.608	(0.406/8.022)
class 4	2.639	2.352	(0.675/6.556)	2.055	(0.441/8.23)	-0.938	(0.370/8.300)	2.355	(0.643/4.718)
class 5	-7.304	-3.709	(1.036/6.232)	-7.158	(0.434/5.833)	1.601	(0.494/4.224)	2.105	(0.423/5.063)
sex	-0.513	0.129	(0.249/0.243)	0.096	(0.25/0.271)	-0.177	(0.329/0.186)	-0.211	(0.327/0.235)
age	-0.044	-0.038	(0.016/0.017)	-0.042	(0.016/0.016)	-0.033	(0.021/0.013)	-0.035	(0.021/0.015)
level 2 vs. level 3									
intercept	0.521	-0.665	(0.484/4.612)	-0.121	(0.495/3.301)	-3.194	(0.641/6.481)	-1.149	(0.625/10.231)
class 1	1.685	2.125	(0.429/4.665)	3.346	(0.457/5.542)	0.079	(0.396/10.963)	3.056	(0.410/11.570)
class 2	-0.428	0.809	(0.345/6.072)	-0.214	(0.345/6.264)	3.394	(0.499/6.771)	1.049	(0.514/10.027)

Table 25: Continued

class 3	**2.204**	**2.497**	(0.341/4.588)	**1.777**	(0.346/3.815)	**5.495**	(0.404/7.393)	**-1.217**	(0.342/12.210)
class 4	**0.546**	**1.368**	(0.363/5.041)	**0.925**	(0.354/5.278)	**3.051**	(0.299/11.903)	**1.083**	(0.499/10.176)
class 5	**-4.833**	**-2.001**	(0.496/5.127)	**-3.038**	(0.553/3.873)	**5.891**	(0.375/6.629)	**2.978**	(0.355/11.230)
sex	**-0.274**	**0.103**	(0.191/0.203)	**0.070**	(0.193/0.213)	**-0.208**	(0.266/0.200)	**-0.199**	(0.265/0.182)
age	**-0.008**	**-0.006**	(0.012/0.013)	**-0.006**	(0.012/0.015)	**0.005**	(0.016/0.010)	**0.007**	(0.016/0.008)

					item 3				
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
					level 1 vs. level 3				
intercept	**-7.445**	**-4.267**	(0.683/7.031)	**-6.275**	(0.623/5.74)	**-3.519**	(0.815/4.738)	**-4.616**	(0.784/5.300)
class 1	**8.520**	**6.414**	(0.61/7.781)	**7.516**	(0.465/5.632)	**-0.200**	(0.451/8.187)	**1.185**	(0.531/8.073)
class 2	**2.959**	**-0.495**	(0.64/12.376)	**2.130**	(0.454/9.158)	**2.019**	(0.654/4.699)	**3.066**	(0.756/5.556)
class 3	**4.838**	**2.073**	(0.565/7.372)	**4.334**	(0.504/6.554)	**1.124**	(0.556/5.370)	**4.498**	(0.434/8.071)
class 4	**7.931**	**4.478**	(0.497/9.631)	**3.845**	(0.598/7.312)	**-2.764**	(0.402/9.547)	**3.083**	(0.716/4.716)
class 5	**3.620**	**-0.937**	(0.814/7.222)	**0.705**	(0.682/6.616)	**1.582**	(0.518/4.664)	**1.226**	(0.455/8.407)
sex	**-0.450**	**-0.045**	(0.245/0.287)	**-0.048**	(0.23/0.281)	**-0.506**	(0.344/0.219)	**-0.535**	(0.341/0.254)
age	**0.007**	**0.003**	(0.015/0.018)	**0.001**	(0.015/0.016)	**0.013**	(0.021/0.015)	**0.011**	(0.021/0.015)
					level 2 vs. level 3				
intercept	**-2.650**	**-1.705**	(0.434/4.563)	**-1.573**	(0.476/2.675)	**-3.860**	(0.6516.373)	**-6.512**	(0.631/8.971)
class 1	**3.046**	**2.206**	(0.447/5.463)	**2.291**	(0.391/2.653)	**-0.056**	(0.405/10.105)	**3.676**	(0.400/11.786)
class 2	**0.766**	**-1.087**	(NaN/8.133)	**-0.027**	(0.425/3.721)	**2.898**	(0.519/6.168)	**5.571**	(0.553/8.974)
class 3	**3.383**	**2.371**	(0.295/4.12)	**2.161**	(0.359/3.131)	**7.104**	(0.396/9.567)	**3.237**	(0.339/12.320)
class 4	**3.271**	**1.246**	(0.301/5.021)	**1.314**	(0.508/3.693)	**-3.146**	(0.287/12.328)	**5.839**	(0.508/9.025)
class 5	**0.783**	**-0.554**	(0.33/4.42)	**-0.435**	(0.388/2.679)	**2.789**	(0.391/6.225)	**2.835**	(0.389/11.994)
sex	**-0.794**	**-0.478**	(0.168/0.167)	**-0.496**	(0.168/0.182)	**-0.508**	(0.273/0.290)	**-0.571**	(0.274/0.305)
age	**0.019**	**0.012**	(0.011/0.011)	**0.011**	(0.011/0.011)	0.017	(0.017/0.007)	**0.018**	(0.017/0.006)

					item 4				
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
					level 1 vs. level 3				
intercept	**-6.883**	**-1.018**	(0.664/2.902)	**-0.351**	(0.677/4.27)	**-0.877**	(0.723/2.212)	**-2.761**	(0.712/4.951)
class 1	**10.946**	**4.980**	(0.615/3.237)	**6.890**	(0.771/5.578)	**0.344**	(0.415/4.018)	**2.321**	(0.473/5.877)
class 2	**7.782**	**1.768**	(0.51/4.709)	**1.865**	(0.502/6.252)	**1.373**	(0.554/2.435)	**3.354**	(0.657/5.392)
class 3	**9.085**	**2.698**	(0.592/3.736)	**1.710**	(0.518/5.186)	**0.269**	(0.445/4.048)	**1.021**	(0.389/8.535)
class 4	**11.649**	**2.769**	(0.609/4.364)	**2.508**	(0.572/6.271)	**-2.700**	(0.413/9.136)	**3.124**	(0.595/5.269)
class 5	**2.957**	**-2.581**	(0.545/3.182)	**-4.056**	(0.57/4.825)	**0.968**	(0.420/2.243)	**1.838**	(0.405/6.401)
sex	**-1.833**	**-0.846**	(0.255/0.304)	**-0.967**	(0.265/0.328)	**-0.750**	(0.308/0.249)	**-0.755**	(0.304/0.208)

Table 25: Continued

age	**0.030**	**0.021**	(0.016/0.016)	**0.022**	(0.017/0.018)	**0.018**	(0.019/0.020)	**0.011**	(0.019/0.014)
				level 2 vs. level 3					
intercept	**-0.261**	**0.224**	(0.595/3.029)	**0.629**	(0.598/3.997)	**2.650**	(0.737/5.416)	**0.728**	(0.719/7.563)
class 1	**1.698**	**1.826**	(0.607/3.344)	**3.416**	(0.732/5.679)	**0.435**	(0.451/8.635)	**1.982**	(0.474/9.487)
class 2	**1.665**	**1.546**	(0.458/5.141)	**1.553**	(0.459/4.993)	**-2.160**	(0.588/5.076)	**0.264**	(0.639/7.712)
class 3	**3.753**	**2.689**	(0.537/3.485)	**2.241**	(0.475/4.993)	**0.890**	(0.470/9.373)	**-2.886**	(0.406/10.318)
class 4	**4.873**	**1.742**	(0.576/4.291)	**1.433**	(0.5/5.438)	**-1.864**	(0.435/10.567)	**0.06**	(0.582/7.890)
class 5	**-2.089**	**-2.616**	(0.409/3.572)	**-3.551**	(0.441/4.477)	**-2.123**	(0.423/5.485)	**1.763**	(0.422/9.511)
sex	**-0.783**	**-0.144**	(0.235/0.271)	**-0.243**	(0.244/0.305)	**0.215**	(0.315/0.128)	**0.198**	(0.312/0.130)
age	**0.008**	**0.000**	(0.015/0.014)	**0.002**	(0.015/0.016)	**-0.013**	(0.019/0.021)	**-0.012**	(0.019/0.015)

				item 5					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				level 1 vs. level 3					
intercept	**0.199**	**-0.146**	(0.554/3.284)	**1.120**	(0.556/3.736)	**-1.728**	(0.720/5.503)	**-0.972**	(0.736/3.701)
class 1	**3.300**	**3.172**	(0.462/3.377)	**2.625**	(0.483/4.781)	**2.047**	(0.507/7.102)	**4.249**	(0.490/5.514)
class 2	**0.106**	**0.084**	(0.456/5.887)	**0.452**	(0.413/5.994)	**2.753**	(0.560/5.944)	**1.871**	(0.649/3.329)
class 3	**1.796**	**1.673**	(0.663/4.025)	**0.502**	(0.396/4.967)	**5.346**	(0.429/6.934)	**-0.393**	(0.417/7.445)
class 4	**0.370**	**-0.057**	(0.407/6.762)	**0.217**	(0.437/6.276)	**3.071**	(0.366/10.917)	**1.958**	(0.613/3.461)
class 5	**-2.095**	**-1.963**	(0.435/3.643)	**-3.409**	(0.458/3.913)	**2.394**	(0.417/5.747)	**4.028**	(0.442/5.164)
sex	**-0.767**	**-0.503**	(0.211/0.237)	**-0.555**	(0.21/0.241)	**-0.456**	(0.301/0.113)	**-0.458**	(0.296/0.082)
age	**0.000**	**0.005**	(0.013/0.013)	**0.004**	(0.013/0.011)	**-0.011**	(0.018/0.014)	**-0.011**	(0.018/0.015)
				level 2 vs. level 3					
intercept	**0.889**	**0.673**	(0.537/4.004)	**1.909**	(0.522/4.74)	**-1.468**	(0.739/5.766)	**5.37**	(0.761/6.028)
class 1	**0.479**	**0.784**	(0.501/3.959)	**-0.230**	(0.514/5.11)	**5.650**	(0.533/7.075)	**-4.847**	(0.514/6.073)
class 2	**-0.379**	**-0.530**	(0.458/6.344)	**-0.877**	(0.379/6.786)	**2.729**	(0.577/5.835)	**-3.92**	(0.643/6.269)
class 3	**20.620**	**2.903**	(0.574/6.413)	**1.225**	(0.378/6.41)	**1.693**	(0.445/6.816)	**-6.652**	(0.452/7.093)
class 4	**-0.073**	**-1.121**	(0.42/7.132)	**-0.640**	(0.427/7.106)	**-0.662**	(0.387/5.926)	**-4.043**	(0.627/6.073)
class 5	**-1.232**	**-1.346**	(0.376/0.203)	**-2.735**	(0.36/4.973)	**2.679**	(0.424/5.926)	**-4.926**	(0.462/5.787)
sex	**-0.586**	**-0.445**	(0.202/0.203)	**-0.486**	(0.202/0.211)	**-0.289**	(0.310/0.219)	**0.316**	(0.310/0.176)
age	**-0.011**	**-0.011**	(0.013/0.013)	**-0.010**	(0.013/0.012)	**-0.028**	(0.019/0.019)	**-0.026**	(0.019/0.020)

Table 26: Average parameters estimations for 100 replication in 6-class model, N=1000
(standard error in multinomial regression / sample standard error for 100 replication)

Class 1 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**0.408**	**-0.378**	(0.243/0.710)	**-0.509**	(0.251/0.815)	**-0.091**	(8.224/15.240)	**0.245**	(1.917/4.187)
occup	**-0.464**	**0.512**	(0.270/0.332)	**0.577**	(0.283/0.386)	**0.222**	(15.559/4.793)	**0.640**	(21.939/4.518)
dprime	**0.092**	**0.375**	(0.074/0.16)	**0.408**	(0.077/0.162)	**-0.043**	(0.778/2.913)	**1.103**	(1.055/1.934)

Class 2 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.313**	**-0.113**	(0.250/1.019)	**-0.059**	(0.249/1.126)	**3.775**	(4.373/10.414)	**0.569**	(1.570/3.573)
occup	**-0.021**	**-0.072**	(0.335/0.435)	**-0.029**	(0.341/0.481)	**1.405**	(7.792/2.818)	**1.532**	(15.992/3.842)
dprime	**0.195**	**0.026**	(0.082/0.206)	**0.053**	(0.083/0.221)	**-0.811**	(0.967/1.886)	**1.076**	(0.964/1.993)

Class 3 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.482**	**0.244**	(0.223/0.724)	**0.217**	(0.227/0.899)	**3.499**	(4.440/9.582)	**1.420**	(1.643/3.995)
occup	**0.149**	**-0.119**	(0.335/0.437)	**-0.046**	(0.301/0.433)	**-0.604**	(7.637/5.987)	**0.878**	(20.498/4.457)
dprime	**0.265**	**0.157**	(0.085/0.237)	**0.186**	(0.074/0.176)	**-10.190**	(1.084/1.731)	**0.896**	(1.003/2.007)

Class 4 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.033**	**-0.334**	(0.262/1.061)	**-0.395**	(0.269/1.039)	**0.532**	(21.239/12.483)	**1.084**	(1.491/3.347)
occup	**-0.187**	**0.079**	(0.335/0.437)	**0.109**	(0.352/0.541)	**2.334**	(21.624/6.257)	**1.884**	(15.843/2.880)
dprime	**0.181**	**0.095**	(0.085/0.237)	**0.124**	(0.088/0.212)	**-0.862**	(1.176/1.964)	**0.906**	(0.946/1.972)

Class 5 vs. Class 6

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-0.171**	**0.688**	(0.208/0.828)	**0.603**	(0.212/0.836)	**4.661**	(4.258/9.314)	**-14.664**	(11.243/40.854)
occup	**0.158**	**-0.052**	(0.291/0.342)	**-0.041**	(0.306/0.395)	**1.448**	(7.663/3.015)	**8.100**	(25.575/14.182)
dprime	**0.104**	**-0.075**	(0.071/0.159)	**-0.055**	(0.073/0.152)	**-0.827**	(0.933/1.807)	**3.083**	(4.979/4.995)

Table 27: Average conditional Probability for 100 replication in 6-class model, N=1000 (sample standard deviance in parentheses)

				Class 1					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.955**	**0.879**	(0.057)	**0.910**	(0.056)	**0.388**	(0.219)	**0.379**	(0.221)
Level 2	**0.044**	**0.111**	(0.053)	**0.084**	(0.051)	**0.450**	(0.310)	**0.429**	(0.290)
Level 3	**0.001**	**0.010**	(0.009)	**0.005**	(0.008)	**0.160**	(0.099)	**0.192**	(0.112)
				item 2					
Level 1	**0.727**	**0.710**	(0.143)	**0.733**	(0.122)	**0.235**	(0.138)	**0.223**	(0.113)
Level 2	**0.235**	**0.232**	(0.115)	**0.222**	(0.099)	**0.291**	(0.181)	**0.455**	(0.269)
Level 3	**0.038**	**0.059**	(0.038)	**0.045**	(0.034)	**0.473**	(0.297)	**0.322**	(0.164)
				item 3					
Level 1	**0.508**	**0.485**	(0.215)	**0.504**	(0.160)	**0.167**	(0.094)	**0.127**	(0.092)
Level 2	**0.318**	**0.343**	(0.150)	**0.345**	(0.118)	**0.211**	(0.130)	**0.291**	(0.167)
Level 3	**0.174**	**0.172**	(0.080)	**0.151**	(0.057)	**0.620**	(0.217)	**0.582**	(0.210)
				item 4					
Level 1	**0.929**	**0.858**	(0.064)	**0.893**	(0.064)	**0.381**	(0.218)	**0.321**	(0.175)
Level 2	**0.056**	**0.120**	(0.058)	**0.094**	(0.058)	**0.429**	(0.320)	**0.460**	(0.285)
Level 3	**0.015**	**0.022**	(0.014)	**0.014**	(0.013)	**0.188**	(0.109)	**0.219**	(0.122)
				item 5					
Level 1	**0.883**	**0.829**	(0.066)	**0.850**	(0.071)	**0.377**	(0.222)	**0.525**	(0.241)
Level 2	**0.078**	**0.119**	(0.052)	**0.103**	(0.057)	**0.461**	(0.308)	**0.265**	(0.134)
Level 3	**0.039**	**0.052**	(0.025)	**0.047**	(0.027)	**0.161**	(0.138)	**0.210**	(0.116)

				Class 2					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.439**	**0.395**	(0.230)	**0.376**	(0.239)	**0.386**	(0.063)	**0.433**	(0.044)
Level 2	**0.384**	**0.430**	(0.181)	**0.441**	(0.215)	**0.371**	(0.088)	**0.337**	(0.079)
Level 3	**0.177**	**0.176**	(0.204)	**0.183**	(0.246)	**0.243**	(0.118)	**0.230**	(0.073)
				item 2					
Level 1	**0.071**	**0.144**	(0.178)	**0.140**	(0.201)	**0.182**	(0.049)	**0.131**	(0.059)
Level 2	**0.397**	**0.461**	(0.202)	**0.413**	(0.193)	**0.455**	(0.058)	**0.440**	(0.099)
Level 3	**0.532**	**0.395**	(0.250)	**0.447**	(0.261)	**0.363**	(0.091)	**0.429**	(0.109)
				item 3					
Level 1	**0.009**	**0.123**	(0.233)	**0.129**	(0.243)	**0.149**	(0.048)	**0.151**	(0.083)

Table 27: Continued

Level 2	**0.156**	**0.248**	(0.177)	**0.230**	(0.172)	**0.297**	(0.103)	**0.294**	(0.097)
Level 3	**0.834**	**0.629**	(0.256)	**0.641**	(0.264)	**0.554**	(0.192)	**0.555**	(0.125)
					item 4				
Level 1	**0.361**	**0.346**	(0.239)	**0.357**	(0.230)	**0.414**	(0.097)	**0.373**	(0.124)
Level 2	**0.498**	**0.459**	(0.217)	**0.458**	(0.196)	**0.315**	(0.086)	**0.416**	(0.103)
Level 3	**0.141**	**0.195**	(0.212)	**0.185**	(0.199)	**0.271**	(0.109)	**0.211**	(0.076)
					item 5				
Level 1	**0.335**	**0.376**	(0.240)	**0.386**	(0.239)	**0.396**	(0.094)	**0.344**	(0.090)
Level 2	**0.306**	**0.325**	(0.197)	**0.300**	(0.187)	**0.330**	(0.062)	**0.385**	(0.050)
Level 3	**0.359**	**0.299**	(0.267)	**0.314**	(0.256)	**0.274**	(0.077)	**0.271**	(0.085)

Class 3

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
					item 1				
Level 1	**0.609**	**0.484**	(0.131)	**0.496**	(0.155)	**0.339**	(0.210)	**0.368**	(0.185)
Level 2	**0.360**	**0.465**	(0.119)	**0.423**	(0.134)	**0.269**	(0.157)	**0.471**	(0.263)
Level 3	**0.031**	**0.051**	(0.091)	**0.081**	(0.186)	**0.392**	(0.362)	**0.161**	(0.086)
					item 2				
Level 1	**0.217**	**0.151**	(0.093)	**0.184**	(0.124)	**0.171**	(0.116)	**0.180**	(0.114)
Level 2	**0.714**	**0.696**	(0.100)	**0.655**	(0.160)	**0.488**	(0.326)	**0.337**	(0.176)
Level 3	**0.069**	**0.152**	(0.098)	**0.161**	(0.169)	**0.381**	(0.246)	**0.483**	(0.266)
					item 3				
Level 1	**0.020**	**0.071**	(0.059)	**0.083**	(0.052)	**0.106**	(0.081)	**0.142**	(0.077)
Level 2	**0.705**	**0.584**	(0.145)	**0.570**	(0.149)	**0.474**	(0.322)	**0.262**	(0.137)
Level 3	**0.275**	**0.346**	(0.151)	**0.347**	(0.167)	**0.419**	(0.277)	**0.596**	(0.209)
					item 4				
Level 1	**0.242**	**0.304**	(0.134)	**0.314**	(0.137)	**0.328**	(0.188)	**0.350**	(0.182)
Level 2	**0.732**	**0.629**	(0.138)	**0.592**	(0.163)	**0.485**	(0.309)	**0.260**	(0.145)
Level 3	**0.026**	**0.067**	(0.107)	**0.094**	(0.176)	**0.187**	(0.157)	**0.390**	(0.300)
					item 5				
Level 1	**0.319**	**0.300**	(0.138)	**0.336**	(0.135)	**0.522**	(0.522)	**0.343**	(0.178)
Level 2	**0.618**	**0.586**	(0.165)	**0.544**	(0.165)	**0.272**	(0.159)	**0.250**	(0.152)
Level 3	**0.063**	**0.114**	(0.085)	**0.120**	(0.140)	**0.206**	(0.142)	**0.407**	(0.293)

Class 4

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	

Table 27: Continued

					item 1				
Level 1	**0.164**	**0.415**	(0.277)	**0.412**	(0.288)	**0.272**	(0.265)	**0.383**	(0.106)
Level 2	**0.811**	**0.479**	(0.262)	**0.445**	(0.273)	**0.443**	(0.361)	**0.374**	(0.086)
Level 3	**0.025**	**0.106**	(0.206)	**0.143**	(0.286)	**0.285**	(0.407)	**0.243**	(0.064)
					item 2				
Level 1	**0.085**	**0.217**	(0.245)	**0.206**	(0.266)	**0.121**	(0.137)	**0.153**	(0.058)
Level 2	**0.607**	**0.487**	(0.195)	**0.468**	(0.226)	**0.476**	(0.355)	**0.439**	(0.104)
Level 3	**0.307**	**0.296**	(0.215)	**0.325**	(0.266)	**0.403**	(0.374)	**0.418**	(0.095)
					item 3				
Level 1	**0.330**	**0.272**	(0.325)	**0.206**	(0.244)	**0.101**	(0.114)	**0.137**	(0.079)
Level 2	**0.466**	**0.310**	(0.187)	**0.352**	(0.192)	**0.178**	(0.166)	**0.347**	(0.102)
Level 3	**0.203**	**0.418**	(0.249)	**0.442**	(0.254)	**0.721**	(0.259)	**0.526**	(0.140)
					item 4				
Level 1	**0.581**	**0.478**	(0.243)	**0.463**	(0.265)	**0.324**	(0.319)	**0.362**	(0.075)
Level 2	**0.414**	**0.378**	(0.191)	**0.376**	(0.215)	**0.397**	(0.374)	**0.391**	(0.041)
Level 3	**0.005**	**0.144**	(0.190)	**0.161**	(0.252)	**0.279**	(0.409)	**0.247**	(0.045)
					item 5				
Level 1	**0.360**	**0.440**	(0.268)	**0.423**	(0.267)	**0.507**	(0.358)	**0.377**	(0.052)
Level 2	**0.343**	**0.274**	(0.190)	**0.285**	(0.194)	**0.204**	(0.192)	**0.353**	(0.107)
Level 3	**0.296**	**0.286**	(0.287)	**0.291**	(0.301)	**0.289**	(0.404)	**0.269**	(0.064)

					Class 5				
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
					item 1				
Level 1	**0.009**	**0.035**	(0.046)	**0.018**	(0.014)	**0.340**	(0.065)	**0.300**	(0.154)
Level 2	**0.019**	**0.134**	(0.075)	**0.093**	(0.053)	**0.384**	(0.032)	**0.470**	(0.266)
Level 3	**0.972**	**0.831**	(0.099)	**0.889**	(0.062)	**0.276**	(0.047)	**0.230**	(0.118)
					item 2				
Level 1	**0.000**	**0.008**	(0.027)	**0.003**	(0.004)	**0.142**	(0.038)	**0.158**	(0.078)
Level 2	**0.009**	**0.087**	(0.046)	**0.067**	(0.036)	**0.379**	(0.019)	**0.479**	(0.258)
Level 3	**0.991**	**0.905**	(0.063)	**0.931**	(0.037)	**0.478**	(0.046)	**0.363**	(0.182)
					item 3				
Level 1	**0.018**	**0.013**	(0.009)	**0.014**	(0.010)	**0.108**	(0.033)	**0.144**	(0.081)
Level 2	**0.157**	**0.130**	(0.040)	**0.135**	(0.029)	**0.287**	(0.060)	**0.193**	(0.127)
Level 3	**0.825**	**0.858**	(0.044)	**0.851**	(0.032)	**0.605**	(0.042)	**0.663**	(0.173)
					item 4				
Level 1	**0.019**	**0.051**	(0..040)	**0.036**	(0.022)	**0.321**	(0.043)	**0.320**	(0.169)

Table 27: Continued

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
Level 2	**0.075**	**0.111**	(0.053)	**0.087**	(0.047)	**0.372**	(0.054)	**0.416**	(0.298)
Level 3	**0.907**	**0.838**	(0.079)	**0.877**	(0.062)	**0.307**	(0.060)	**0.264**	(0.144)
item 5									
Level 1	**0.070**	**0.092**	(0.050)	**0.086**	(0.040)	**0.322**	(0.030)	**0.496**	(0.252)
Level 2	**0.248**	**0.235**	(0.084)	**0.221**	(0.076)	**0.352**	(0.031)	**0.278**	(0.156)
Level 3	**0.682**	**0.673**	(0.115)	**0.694**	(0.104)	**0.326**	(0.056)	**0.225**	(0.115)

				Class 6					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
item 1									
Level 1	**0.002**	**0.248**	(0.231)	**0.236**	(0.220)	**0.253**	(0.171)	**0.258**	(0.203)
Level 2	**0.709**	**0.465**	(0.205)	**0.508**	(0.232)	**0.537**	(0.268)	**0.432**	(0.322)
Level 3	**0.289**	**0.288**	(0.253)	**0.256**	(0.269)	**0.210**	(0.122)	**0.310**	(0.355)
item 2									
Level 1	**0.009**	**0.109**	(0.225)	**0.098**	(0.181)	**0.157**	(0.103)	**0.140**	(0.114)
Level 2	**0.529**	**0.405**	(0.206)	**0.437**	(0.189)	**0.298**	(0.182)	**0.442**	(0.320)
Level 3	**0.462**	**0.487**	(0.250)	**0.465**	(0.240)	**0.545**	(0.269)	**0.418**	(0.323)
item 3									
Level 1	**0.001**	**0.098**	(0.232)	**0.071**	(0.185)	**0.123**	(0.070)	**0.113**	(0.088)
Level 2	**0.080**	**0.209**	(0.158)	**0.206**	(0.164)	**0.219**	(0.125)	**0.217**	(0.182)
Level 3	**0.919**	**0.692**	(0.244)	**0.723**	(0.239)	**0.658**	(0.192)	**0.670**	(0.267)
item 4									
Level 1	**0.001**	**0.245**	(0.218)	**0.221**	(0.213)	**0.287**	(0.162)	**0.256**	(0.199)
Level 2	**0.400**	**0.390**	(0.191)	**0.403**	(0.224)	**0.447**	(0.312)	**0.415**	(0.328)
Level 3	**0.600**	**0.364**	(0.249)	**0.376**	(0.281)	**0.266**	(0.151)	**0.329**	(0.345)
item 5									
Level 1	**0.272**	**0.322**	(0.194)	**0.306**	(0.185)	**0.291**	(0.168)	**0.277**	(0.216)
Level 2	**0.404**	**0.386**	(0.198)	**0.423**	(0.200)	**0.267**	(0.159)	**0.547**	(0.357)
Level 3	**0.324**	**0.292**	(0.244)	**0.271**	(0.236)	**0.442**	(0.313)	**0.175**	(0.164)

Table 28: Average Latent Prevalences for 100 replication in 6-class model, N=1000 (sample standard deviance in parentheses)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.200**	**0.253**	(0.074)	**0.237**	(0.080)	**0.162**	(0.185)	**0.130**	(0.118)
class2	**0.200**	**0.114**	(0.046)	**0.128**	(0.064)	**0.164**	(0.156)	**0.107**	(0.078)
class3	**0.150**	**0.208**	(0.064)	**0.220**	(0.080)	**0.149**	(0.165)	**0.270**	(0.191)
class4	**0.150**	**0.117**	(0.060)	**0.121**	(0.073)	**0.061**	(0.061)	**0.106**	(0.031)
class5	**0.150**	**0.189**	(0.054)	**0.174**	(0.046)	**0.288**	(0.047)	**0.198**	(0.152)
class6	**0.150**	**0.120**	(0.046)	**0.120**	(0.056)	**0.176**	(0.157)	**0.189**	(0.195)

Table 29: Average Correlation Coefficients for 100 replication in 6-class model, N=1000 (total number of not NA values in parentheses)

	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.098**	(100)	**0.114**	(100)	**0.258**	(94)	**0.331**	(95)
class2	**0.124**	(100)	**0.123**	(100)	**0.290**	(99)	**0.282**	(100)
class3	**0.096**	(100)	**0.097**	(100)	**0.266**	(95)	**0.277**	(99)
class4	**0.119**	(100)	**0.121**	(100)	**0.244**	(88)	**0.285**	(95)
class5	**0.194**	(100)	**0.170**	(100)	**0.297**	(96)	**0.313**	(86)
class6	**0.148**	(100)	**0.130**	(100)	**0.306**	(96)	**0.256**	(99)

Table 30: Average Match Proportions for 100 replication in 6-class model, N=1000

	K_Corr	K_Cova	D_Corr	D_Cova
class1	**0.785**	**0.495**	**0.233**	**0.163**
class2	**0.205**	**0.225**	**0.2**	**0.152**
class3	**0.533**	**0.640**	**0.182**	**0.329**
class4	**0.333**	**0.366**	**0.093**	**0.155**
class5	**0.886**	**0.526**	**0.391**	**0.251**
class6	**0.353**	**0.360**	**0.257**	**0.266**

Table 31: Average parameters estimations for 100 replication in 2-class model, N=150
(standard error in multinomial regression / sample standard error for 100 replication)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
item 1									
level 1 vs. level 3									
intercept	**-3.743**	**-3.288**	(1.290/2.053)	**-3.144**	(1.560/2.493)	**-3.288**	(1.290/2.053)	**-3.144**	(1.560/2.493)
class 1	**6.970**	**5.424**	(2.273/3.256)	**6.753**	(1.119/3.854)	**5.424**	(2.273/3.256)	**6.753**	(1.119/3.854)
sex	**0.391**	**0.283**	(0.572/1.150)	**0.438**	(0.59/0.587)	**0.283**	(0.572/1.150)	**0.438**	(0.590/0.587)
age	**0.051**	**0.050**	(0.037/0.033)	**0.025**	(0.036/0.037)	**0.050**	(0.037/0.033)	**0.025**	(0.036/0.037)
level 2 vs. level 3									
intercept	**-0.866**	**-0.944**	(1.115/1.798)	**-0.915**	(1.180/1.819)	**-0.944**	(1.115/1.798)	**-0.915**	(1.180/1.819)
class 1	**3.962**	**3.199**	(2.217/3.187)	**4.161**	(0.847/3.717)	**3.199**	(2.217/3.187)	**4.161**	(0.847/3.717)
sex	**0.501**	**0.416**	(0.522/1.070)	**0.535**	(0.525/0.508)	**0.416**	(0.522/1.070)	**0.535**	(0.525/0.508)
age	**0.012**	**0.014**	(0.035/0.036)	**0.005**	(0.032/0.031)	**0.014**	(0.035/0.036)	**0.005**	(0.032/0.031)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
item 2									
level 1 vs. level 3									
intercept	**-3.986**	**-4.514**	(2.073/3.886)	**-5.861**	(3.782/14.607)	**-4.514**	(2.073/3.886)	**-5.861**	(3.782/14.607)
class 1	**6.047**	**5.989**	(1.644/4.154)	**8.354**	(3.688/4.391)	**5.989**	(1.644/4.154)	**8.354**	(3.688/4.391)
sex	**0.311**	**0.371**	(0.555/0.573)	**0.358**	(0.571/0.625)	**0.371**	(0.555/0.573)	**0.358**	(0.571/0.625)
age	**-0.014**	**0.006**	(0.034/0.032)	**-0.033**	(0.034/0.035)	**0.006**	(0.034/0.032)	**-0.033**	(0.034/0.035)
level 2 vs. level 3									
intercept	**-0.923**	**-1.297**	(0.980/1.842)	**-0.758**	(1.047/1.113)	**-1.297**	(0.980/1.842)	**-0.758**	(1.047/1.113)
class 1	**2.586**	**2.466**	(0.525/1.842)	**2.665**	(0.713/1.141)	**2.466**	(0.525/1.842)	**2.665**	(0.713/1.141)
sex	**0.312**	**0.365**	(0.464/0.481)	**0.375**	(0.468/0.553)	**0.365**	(0.464/0.481)	**0.375**	(0.468/0.553)
age	**-0.002**	**0.008**	(0.029/0.026)	**-0.012**	(0.028/0.027)	**0.008**	(0.029/0.026)	**-0.012**	(0.028/0.027)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
item 3									
level 1 vs. level 3									
intercept	**-4.839**	**-5.929**	(3.689/4.113)	**-7.544**	(5.633/4.615)	**-5.929**	(3.689/4.113)	**-7.544**	(5.633/4.615)
class 1	**4.373**	**5.848**	(3.379/4.430)	**7.505**	(5.33/4.309)	**5.848**	(3.379/4.430)	**7.505**	(5.33/4.309)
sex	**-0.053**	**-0.133**	(0.802/0.561)	**-0.088**	(0.541/0.543)	**-0.133**	(0.802/0.561)	**-0.088**	(0.541/0.543)
age	**0.024**	**0.026**	(0.040/0.037)	**0.014**	(0.032/0.037)	**0.026**	(0.040/0.037)	**0.014**	(0.032/0.037)
level 2 vs. level 3									
intercept	**-1.780**	**-1.911**	(1.086/1.497)	**-1.752**	(0.989/1.022)	**-1.911**	(1.086/1.497)	**-1.752**	(0.989/1.022)

Table 31: Continued

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
class 1	**1.990**	**1.875**	(0.612/1.634)	**1.963**	(0.462/0.977)	**1.875**	(0.612/1.634)	**1.963**	(0.462/0.977)
sex	**-0.442**	**-0.522**	(0.427/0.359)	**-0.419**	(0.424/0.447)	**-0.522**	(0.427/0.359)	**-0.419**	(0.424/0.447)
age	**0.019**	**0.022**	(0.026/0.030)	**0.013**	(0.025/0.025)	**0.022**	(0.026/0.030)	**0.013**	(0.025/0.025)

				item 4					
				level 1 vs. level 3					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-2.893**	**-2.576**	(1.251/1.573)	**-2.417**	(1.358/2.062)	**-2.576**	(1.251/1.573)	**-2.417**	(1.358/2.062)
class 1	**5.733**	**5.484**	(1.645/3.440)	**6.018**	(1.278/3.560)	**5.484**	(1.645/3.440)	**6.018**	(1.278/3.560)
sex	**-0.462**	**-0.432**	(0.565/0.611)	**-0.375**	(0.569/0.649)	**-0.432**	(0.565/0.611)	**-0.375**	(0.569/0.649)
age	**0.042**	**0.043**	(0.036/0.028)	**0.022**	(0.035/0.041)	**0.043**	(0.036/0.028)	**0.022**	(0.035/0.041)
				level 2 vs. level 3					
intercept	**-0.727**	**-0.640**	(1.092/1.151)	**-0.549**	(1.163/1.778)	**-0.640**	(1.092/1.151)	**-0.549**	(1.163/1.778)
class 1	**3.630**	**3.530**	(1.580/3.071)	**4.134**	(1.174/3.586)	**3.530**	(1.580/3.071)	**4.134**	(1.174/3.586)
sex	**0.105**	**0.201**	(0.514/0.567)	**0.207**	(0.513/0.537)	**0.201**	(0.514/0.567)	**0.207**	(0.513/0.537)
age	**0.007**	**0.006**	(0.033/0.032)	**-0.006**	(0.032/0.037)	**0.006**	(0.033/0.032)	**-0.006**	(0.032/0.037)

				item 5					
				level 1 vs. level 3					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-1.218**	**-1.715**	(1.071/2.433)	**-1.211**	(1.125/1.698)	**-1.715**	(1.071/2.433)	**-1.211**	(1.125/1.698)
class 1	**2.900**	**3.279**	(0.597/2.463)	**3.051**	(0.555/1.591)	**3.279**	(0.597/2.463)	**3.051**	(0.555/1.591)
sex	**-0.268**	**-0.262**	(0.505/0.509)	**-0.301**	(0.484/0.487)	**-0.262**	(0.505/0.509)	**-0.301**	(0.484/0.487)
age	**0.011**	**0.016**	(0.031/0.030)	**0.006**	(0.029/0.03)	**0.016**	(0.031/0.030)	**0.006**	(0.029/0.03)
				level 2 vs. level 3					
intercept	**0.373**	**-0.098**	(1.010/2.451)	**0.332**	(1.072/1.102)	**-0.098**	(1.010/2.451)	**0.332**	(1.072/1.102)
class 1	**1.223**	**1.771**	(0.573/2.800)	**1.455**	(0.518/1.245)	**1.771**	(0.573/2.800)	**1.455**	(0.518/1.245)
sex	**-0.292**	**-0.346**	(0.493/0.477)	**-0.309**	(0.472/0.517)	**-0.346**	(0.493/0.477)	**-0.309**	(0.472/0.517)
age	**-0.013**	**-0.013**	(0.031/0.029)	**-0.015**	(0.029/0.03)	**-0.013**	(0.031/0.029)	**-0.015**	(0.029/0.03)

Table 32: Average parameters estimations for 100 replication in 2-class model, N=150 (standard error in multinomial regression / sample standard error for 100 replication)

				Class 1 vs. Class 2					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**0.577**	**0.424**	(0.318/0.666)	**-0.439**	(0.337/0.635)	**0.424**	(0.318/0.666)	**-0.439**	(0.337/0.635)
occup	**0.307**	**0.253**	(0.389/0.284)	**0.306**	(0.407/0.438)	**0.253**	(0.389/0.284)	**0.306**	(0.407/0.438)
dprime	**-0.103**	**-0.079**	(0.098/0.072)	**0.309**	(0.108/0.120)	**-0.079**	(0.098/0.072)	**0.309**	(0.108/0.120)

Table 33: Average conditional Probability for 100 replication in 2-class model, N=150 (sample standard deviance in parentheses)

			Class 1						
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.765**	**0.709**	(0.074)	**0.696**	(0.087)	**0.709**	(0.074)	**0.696**	(0.087)
Level 2	**0.229**	**0.251**	0.059	**0.266**	(0.070)	**0.251**	0.059	**0.266**	(0.070)
Level 3	**0.005**	**0.039**	0.04	**0.038**	(0.038)	**0.039**	0.04	**0.038**	(0.038)
				item2					
Level 1	**0.457**	**0.440**	0.101	**0.418**	(0.075)	**0.440**	0.101	**0.418**	(0.075)
Level 2	**0.464**	**0.455**	0.087	**0.465**	(0.053)	**0.455**	0.087	**0.465**	(0.053)
Level 3	**0.079**	**0.105**	0.064	**0.117**	(0.065)	**0.105**	0.064	**0.117**	(0.065)
				item3					
Level 1	**0.325**	**0.336**	0.149	**0.311**	(0.114)	**0.336**	0.149	**0.311**	(0.114)
Level 2	**0.441**	**0.417**	0.107	**0.423**	(0.084)	**0.417**	0.107	**0.423**	(0.084)
Level 3	**0.234**	**0.247**	0.086	**0.266**	(0.082)	**0.247**	0.086	**0.266**	(0.082)
				item4					
Level 1	**0.683**	**0.630**	0.084	**0.620**	(0.077)	**0.630**	0.084	**0.620**	(0.077)
Level 2	**0.304**	**0.328**	0.069	**0.331**	(0.057)	**0.328**	0.069	**0.331**	(0.057)
Level 3	**0.012**	**0.042**	0.044	**0.049**	(0.050)	**0.042**	0.044	**0.049**	(0.050)
				item5					
Level 1	**0.647**	**0.631**	0.07	**0.601**	(0.074)	**0.631**	0.07	**0.601**	(0.074)
Level 2	**0.258**	**0.267**	0.064	**0.282**	(0.057)	**0.267**	0.064	**0.282**	(0.057)
Level 3	**0.095**	**0.102**	0.04	**0.118**	(0.052)	**0.102**	0.04	**0.118**	(0.052)

			Class 2						
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.082**	**0.163**	0.107	**0.114**	(0.086)	**0.163**	0.107	**0.114**	(0.086)
Level 2	**0.415**	**0.390**	0.086	**0.362**	(0.093)	**0.390**	0.086	**0.362**	(0.093)
Level 3	**0.504**	**0.446**	0.106	**0.524**	(0.126)	**0.446**	0.106	**0.524**	(0.126)
				item2					
Level 1	**0.009**	**0.059**	0.062	**0.026**	(0.051)	**0.059**	0.062	**0.026**	(0.051)
Level 2	**0.304**	**0.316**	0.084	**0.282**	(0.076)	**0.316**	0.084	**0.282**	(0.076)
Level 3	**0.687**	**0.625**	0.114	**0.692**	(0.107)	**0.625**	0.114	**0.692**	(0.107)
				item 3					
Level 1	**0.014**	**0.036**	0.040	**0.018**	(0.024)	**0.036**	0.040	**0.018**	(0.024)

Table 33: Continued

Level 2	**0.202**	**0.205**	0.085	**0.196**	(0.080)	**0.205**	0.085	**0.196**	(0.080)
Level 3	**0.784**	**0.759**	0.104	**0.786**	(0.091)	**0.759**	0.104	**0.786**	(0.091)
item4									
Level 1	**0.098**	**0.151**	0.088	**0.117**	(0.076)	**0.151**	0.088	**0.117**	(0.076)
Level 2	**0.356**	**0.360**	0.071	**0.326**	(0.087)	**0.360**	0.071	**0.326**	(0.087)
Level 3	**0.547**	**0.489**	0.107	**0.557**	(0.124)	**0.489**	0.107	**0.557**	(0.124)
item5									
Level 1	**0.173**	**0.193**	0.106	**0.178**	(0.081)	**0.193**	0.106	**0.178**	(0.081)
Level 2	**0.368**	**0.350**	0.107	**0.345**	(0.073)	**0.350**	0.107	**0.345**	(0.073)
Level 3	**0.459**	**0.457**	0.147	**0.477**	(0.101)	**0.457**	0.147	**0.477**	(0.101)

Table 34: Average Latent Prevalences for 100 replication in 2-class model, N=150 (sample standard deviance in parentheses)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.600**	**0.574**	(0.138)	**0.541**	(0.114)	**0.574**	(0.138)	**0.541**	(0.114)
class2	**0.400**	**0.426**	(0.138)	**0.459**	(0.114)	**0.426**	(0.138)	**0.459**	(0.114)

Table 35: Average Correlation Coefficients for 100 replication in 2-class model, N=150 (total number of not NA values in parentheses)

	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.111**	(100)	**0.148**	(100)	**0.111**	(100)	**0.148**	(100)
class2	**0.192**	(100)	**0.190**	(100)	**0.192**	(100)	**0.190**	(100)

Table 36: Average Match Proportions for 100 replication in 2-class model, N=150

	K_Corr	K_Cova	D_Corr	D_Cova
class1	**0.849**	**0.835**	**0.849**	**0.835**
class2	**0.839**	**0.833**	**0.839**	**0.833**

Table 37: Average parameters estimations for 100 replication in 2-class model, N=700
(standard error in multinomial regression / sample standard error for 100 replication)

item 1

level 1 vs. level 3

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-3.743**	**-2.461**	(0.637/1.196)	**-2.905**	(0.653/0.648)	**-2.461**	(0.637/1.196)	**-2.905**	(0.653/0.648)
class 1	**6.970**	**5.136**	(0.453/2.171)	**5.576**	(0.463/1.361)	**5.136**	(0.453/2.171)	**5.576**	(0.463/1.361)
sex	**0.391**	**0.300**	(0.264/0.279)	**0.483**	(0.272/0.278)	**0.300**	(0.264/0.279)	**0.483**	(0.272/0.278)
age	**0.051**	**0.028**	(0.016/0.017)	**0.027**	(0.017/0.017)	**0.028**	(0.016/0.017)	**0.027**	(0.017/0.017)
level 2 vs. level 3									
intercept	**-0.866**	**-0.377**	(0.534/0.645)	**-0.537**	(0.519/0.560)	**-0.377**	(0.534/0.645)	**-0.537**	(0.519/0.560)
class 1	**3.962**	**2.839**	(0.423/1.817)	**3.015**	(0.430/1.273)	**2.839**	(0.423/1.817)	**3.015**	(0.430/1.273)
sex	**0.501**	**0.443**	(0.234/0.238)	**0.572**	(0.230/0.229)	**0.443**	(0.234/0.238)	**0.572**	(0.230/0.229)
age	**0.012**	**0.000**	(0.015/0.015)	**0.000**	(0.015/0.017)	**0.000**	(0.015/0.015)	**0.000**	(0.015/0.017)

item 2

level 1 vs. level 3

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-3.986**	**-3.113**	(0.680/2.865)	**-4.007**	(0.853/2.450)	**-3.113**	(0.680/2.865)	**-4.007**	(0.853/2.450)
class 1	**6.047**	**5.380**	(0.490/3.004)	**6.260**	(0.664/2.353)	**5.380**	(0.490/3.004)	**6.260**	(0.664/2.353)
sex	**0.311**	**0.223**	(0.252/0.282)	**0.364**	(0.265/0.275)	**0.223**	(0.252/0.282)	**0.364**	(0.265/0.275)
age	**-0.014**	**-0.026**	(0.016/0.014)	**-0.027**	(0.017/0.018)	**-0.026**	(0.016/0.014)	**-0.027**	(0.017/0.018)
level 2 vs. level 3									
intercept	**-0.923**	**-0.983**	(0.472/2.365)	**-0.768**	(0.477/0.456)	**-0.983**	(0.472/2.365)	**-0.768**	(0.477/0.456)
class 1	**2.586**	**2.710**	(0.246/2.409)	**2.578**	(0.229/0.404)	**2.710**	(0.246/2.409)	**2.578**	(0.229/0.404)
sex	**0.312**	**0.275**	(0.211/0.218)	**0.358**	(0.212/0.225)	**0.275**	(0.211/0.218)	**0.358**	(0.212/0.225)
age	**-0.002**	**-0.011**	(0.013/0.013)	**-0.010**	(0.013/0.012)	**-0.011**	(0.013/0.013)	**-0.010**	(0.013/0.012)

item 3

level 1 vs. level 3

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-4.839**	**-4.214**	(0.887/1.509)	**-5.492**	(0.895/2.712)	**-4.214**	(0.887/1.509)	**-5.492**	(0.895/2.712)
class 1	**4.373**	**4.622**	(0.639/3.252)	**5.178**	(0.706/2.650)	**4.622**	(0.639/3.252)	**5.178**	(0.706/2.650)
sex	**-0.053**	**-0.043**	(0.247/0.222)	**-0.010**	(0.246/0.251)	**-0.043**	(0.247/0.222)	**-0.010**	(0.246/0.251)
age	**0.024**	**0.013**	(0.015/0.018)	**0.015**	(0.016/0.015)	**0.013**	(0.015/0.018)	**0.015**	(0.016/0.015)
level 2 vs. level 3									
intercept	**-1.780**	**-1.586**	(0.546/1.233)	**-1.707**	(0.450/0.443)	**-1.586**	(0.546/1.233)	**-1.707**	(0.450/0.443)

Table 37: Continued

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
class 1	**1.990**	**1.733**	(0.301/1.536)	**2.018**	(0.199/0.267)	**1.733**	(0.301/1.536)	**2.018**	(0.199/0.267)
sex	**-0.442**	**-0.459**	(0.189/0.186)	**-0.405**	(0.193/0.192)	**-0.459**	(0.189/0.186)	**-0.405**	(0.193/0.192)
age	**0.019**	**0.012**	(0.012/0.012)	**0.013**	(0.012/0.012)	**0.012**	(0.012/0.012)	**0.013**	(0.012/0.012)

item 4

level 1 vs. level 3

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-2.893**	**-2.003**	(0.603/1.790)	**-2.381**	(0.618/0.615)	**-2.003**	(0.603/1.790)	**-2.381**	(0.618/0.615)
class 1	**5.733**	**4.800**	(0.420/2.043)	**5.050**	(0.406/0.766)	**4.800**	(0.420/2.043)	**5.050**	(0.406/0.766)
sex	**-0.462**	**-0.484**	(0.257/0.281)	**-0.396**	(0.262/0.260)	**-0.484**	(0.257/0.281)	**-0.396**	(0.262/0.260)
age	**0.042**	**0.022**	(0.016/0.017)	**0.024**	(0.017/0.016)	**0.022**	(0.016/0.017)	**0.024**	(0.017/0.016)
				level 2 vs. level 3					
intercept	**-0.727**	**-0.343**	(0.524/1.451)	**-0.481**	(0.514/0.495)	**-0.343**	(0.524/1.451)	**-0.481**	(0.514/0.495)
class 1	**3.630**	**3.047**	(0.391/1.768)	**3.071**	(0.372/0.575)	**3.047**	(0.391/1.768)	**3.071**	(0.372/0.575)
sex	**0.105**	**0.094**	(0.232/0.212)	**0.118**	(0.228/0.217)	**0.094**	(0.232/0.212)	**0.118**	(0.228/0.217)
age	**0.007**	**-0.004**	(0.014/0.013)	**-0.002**	(0.015/0.014)	**-0.004**	(0.014/0.013)	**-0.002**	(0.015/0.014)

item 5

level 1 vs. level 3

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**-1.218**	**-1.243**	(0.550/2.446)	**-1.048**	(0.512/0.57)	**-1.243**	(0.550/2.446)	**-1.048**	(0.512/0.570)
class 1	**2.900**	**3.024**	(0.302/2.433)	**2.757**	(0.237/0.713)	**3.024**	(0.302/2.433)	**2.757**	(0.237/0.713)
sex	**-0.268**	**-0.304**	(0.220/0.220)	**-0.239**	(0.219/0.225)	**-0.304**	(0.220/0.220)	**-0.239**	(0.219/0.225)
age	**0.011**	**0.005**	(0.013/0.013)	**0.005**	(0.014/0.014)	**0.005**	(0.013/0.013)	**0.005**	(0.014/0.014)
				level 2 vs. level 3					
intercept	**0.373**	**0.278**	(0.530/2.600)	**0.413**	(0.476/0.460)	**0.278**	(0.530/2.600)	**0.413**	(0.476/0.460)
class 1	**1.223**	**1.436**	(0.295/2.702)	**1.286**	(0.223/0.965)	**1.436**	(0.295/2.702)	**1.286**	(0.223/0.965)
sex	**-0.292**	**-0.331**	(0.216/0.221)	**-0.257**	(0.210/0.203)	**-0.331**	(0.216/0.221)	**-0.257**	(0.210/0.203)
age	**-0.013**	**-0.017**	(0.013/0.013)	**-0.016**	(0.013/0.013)	**-0.017**	(0.013/0.013)	**-0.016**	(0.013/0.013)

Table 38: Average parameters estimations for 100 replication in 2-class model, N=700 (standard error in multinomial regression / sample standard error for 100 replication)

Class 1 vs. Class 2

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
intercept	**0.577**	**0.507**	(0.152/0.730)	**-0.640**	(0.149/0.266)	**0.507**	(0.152/0.730)	**-0.640**	(0.149/0.266)
occup	**0.307**	**-0.072**	(0.183/0.115)	**0.362**	(0.184/0.185)	**-0.072**	(0.183/0.115)	**0.362**	(0.184/0.185)
dprime	**-0.103**	**-0.046**	(0.047/0.031)	**0.338**	(0.048/0.061)	**-0.046**	(0.047/0.031)	**0.338**	(0.048/0.061)

Table 39: Average conditional Probability for 100 replication in 2-class model, N=700 (sample standard deviance in parentheses)

				Class 1					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.765**	**0.708**	(0.080)	**0.717**	(0.036)	**0.708**	(0.080)	**0.717**	(0.036)
Level 2	**0.229**	**0.256**	(0.042)	**0.259**	(0.028)	**0.256**	(0.042)	**0.259**	(0.028)
Level 3	**0.005**	**0.036**	(0.048)	**0.023**	(0.018)	**0.036**	(0.048)	**0.023**	(0.018)
				item2					
Level 1	**0.457**	**0.431**	(0.054)	**0.434**	(0.038)	**0.431**	(0.054)	**0.434**	(0.038)
Level 2	**0.464**	**0.460**	(0.037)	**0.471**	(0.027)	**0.460**	(0.037)	**0.471**	(0.027)
Level 3	**0.079**	**0.109**	(0.067)	**0.096**	(0.031)	**0.109**	(0.067)	**0.096**	(0.031)
				item3					
Level 1	**0.325**	**0.335**	(0.159)	**0.307**	(0.027)	**0.335**	(0.159)	**0.307**	(0.027)
Level 2	**0.441**	**0.408**	(0.101)	**0.436**	(0.027)	**0.408**	(0.101)	**0.436**	(0.027)
Level 3	**0.234**	**0.257**	(0.087)	**0.257**	(0.030)	**0.257**	(0.087)	**0.257**	(0.030)
				item4					
Level 1	**0.683**	**0.630**	(0.061)	**0.639**	(0.042)	**0.630**	(0.061)	**0.639**	(0.042)
Level 2	**0.304**	**0.328**	(0.026)	**0.331**	(0.029)	**0.328**	(0.026)	**0.331**	(0.029)
Level 3	**0.012**	**0.041**	(0.050)	**0.030**	(0.019)	**0.041**	(0.050)	**0.030**	(0.019)
				item5					
Level 1	**0.647**	**0.615**	(0.064)	**0.614**	(0.070)	**0.615**	(0.064)	**0.614**	(0.070)
Level 2	**0.258**	**0.275**	(0.041)	**0.280**	(0.076)	**0.275**	(0.041)	**0.280**	(0.076)
Level 3	**0.095**	**0.110**	(0.040)	**0.106**	(0.023)	**0.110**	(0.040)	**0.106**	(0.023)

				Class 2					
	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
				item 1					
Level 1	**0.082**	**0.142**	(0.117)	**0.093**	(0.044)	**0.142**	(0.117)	**0.093**	(0.044)
Level 2	**0.415**	**0.399**	(0.048)	**0.397**	(0.031)	**0.399**	(0.048)	**0.397**	(0.031)
Level 3	**0.504**	**0.459**	(0.097)	**0.510**	(0.043)	**0.459**	(0.097)	**0.510**	(0.043)
				item2					
Level 1	**0.009**	**0.048**	(0.078)	**0.015**	(0.024)	**0.048**	(0.078)	**0.015**	(0.024)
Level 2	**0.304**	**0.302**	(0.072)	**0.284**	(0.033)	**0.302**	(0.072)	**0.284**	(0.033)
Level 3	**0.687**	**0.650**	(0.131)	**0.701**	(0.048)	**0.650**	(0.131)	**0.701**	(0.048)
				item3					
Level 1	**0.014**	**0.029**	(0.041)	**0.014**	(0.020)	**0.029**	(0.041)	**0.014**	(0.020)

Table 39: Continued

Level 2	**0.202**	**0.218**	(0.075)	**0.188**	(0.027)	**0.218**	(0.075)	**0.188**	(0.027)
Level 3	**0.784**	**0.752**	(0.103)	**0.797**	(0.040)	**0.752**	(0.103)	**0.797**	(0.040)
item4									
Level 1	**0.098**	**0.146**	(0.101)	**0.102**	(0.039)	**0.146**	(0.101)	**0.102**	(0.039)
Level 2	**0.356**	**0.350**	(0.058)	**0.341**	(0.029)	**0.350**	(0.058)	**0.341**	(0.029)
Level 3	**0.547**	**0.504**	(0.116)	**0.557**	(0.044)	**0.504**	(0.116)	**0.557**	(0.044)
item5									
Level 1	**0.173**	**0.189**	(0.090)	**0.172**	(0.042)	**0.189**	(0.090)	**0.172**	(0.042)
Level 2	**0.368**	**0.364**	(0.101)	**0.367**	(0.034)	**0.364**	(0.101)	**0.367**	(0.034)
Level 3	**0.459**	**0.448**	(0.129)	**0.462**	(0.039)	**0.448**	(0.129)	**0.462**	(0.039)

Table 40: Average Latent Prevalences for 100 replication in 2-class model, N=700 (sample standard deviance in parentheses)

	TRUE	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.600**	**0.589**	(0.153)	**0.573**	(0.058)	**0.589**	(0.153)	**0.573**	(0.058)
class2	**0.400**	**0.411**	(0.153)	**0.427**	(0.058)	**0.411**	(0.153)	**0.427**	(0.058)

Table 41: Average Correlation Coefficients for 100 replication in 2-class model, N=700 (total number of not NA values in parentheses)

	K_Corr		K_Cova		D_Corr		D_Cova	
class1	**0.077**	(100)	**0.089**	(100)	**0.077**	(100)	**0.089**	(100)
class2	**0.147**	(100)	**0.143**	(100)	**0.147**	(100)	**0.143**	(100)

Table 42: Average Match Proportions for 100 replication in 2 class model, N=700

	K_Corr	K_Cova	D_Corr	D_Cova
class1	**0.876**	**0.867**	**0.876**	**0.867**
class2	**0.842**	**0.835**	**0.842**	**0.835**

Table 43: Composition of classes of patients at the acute state by the four-class RLCA model with divisive hierarchical clustering method

Factor	Subtype	Class 1	Class 2	Class 3	Class 4
F1	Symptom	**8.67±2.30**	**8.76±2.49**	**8.38±3.18**	**8.3±2.55**
N5	Difficulty in abstract thinking	3.75±1.69	3.90±1.77	3.71±2.00	3.7±1.83
G12	Lack of judgment and insight	4.91±1.21	4.85±1.37	4.66±1.74	4.6±1.27
F2		**19.17±8.62**	**20.93±9.43**	**21.33±9.67**	**20.35±10.5**
N1	Blunted affect	3.04±1.44	3.16±1.59	3.38±1.91	3.5±1.67
N2	Emotional withdrawal	3.17±1.60	3.61±1.72	3.85±1.79	3.45±1.79
N3	Poor rapport	2.82±1.67	2.88±1.86	2.81±2.04	2.75±1.74
N4	Passive/apathetic social withdrawal	3.07±1.54	3.5±1.82	3.47±2.01	3.25±1.94
N6	Lack of spontaneity/flow of conversation	2.64±1.55	3.02±1.88	3.04±2.13	2.6±1.93
G7	Motor retardation	2.08±1.37	2.02±1.23	2.28±1.70	2.25±1.37
G13	Disturbance of volition	2.33±1.34	2.71±1.45	12.47±1.60	2.55±1.63
F3		**15.24±6.59**	**15.71±6.32**	**17.48±8.66**	**16.05±8.88**
P2	Conceptual disorganization	3.19±1.53	3.31±1.58	3.61±2.22	3.45±1.87
N7	Stereotyped thinking	2.74±1.55	2.54±1.61	3.04±1.85	2.7±1.86
G5	Mannerisms and posturing	1.82±1.45	2.02±1.47	2.23±1.81	1.95±1.76
G10	Disorientation	1.83±1.17	1.73±1.12	2.04±1.59	2.25±1.58
G11	Poor attention	2.48±1.40	2.88±1.53	2.90±1.89	2.85±1.78
G15	Preoccupation	3.15±1.67	3.21±1.76	3.61±1.83	2.85±1.95
F4		**10.53±4.26**	**12.29±4.70**	**10.81±4.86**	**11.1±4.32**
G1	Somatic concern	2.47±1.48	2.54±1.59	2.47±1.75	2.25±1.48
G2	Anxiety	2.55±1.44	3.11±1.45	2.42±1.28	2.8±1.32
G3	Guilt feelings	1.40±0.91	1.78±1.22	1.57±1.12	1.4±0.82
G4	Tension	2.091.29	2.64±1.59	2±1.14	2.05±1.09
G6	Depression	2±1.22	2.19±1.29	2.33±1.15	2.6±1.31
F5		**32.52±10.67**	**33.38±10.88**	**37.43±9.75**	**32.3±10.38**
P1	Delusions	5.10±1.40	4.88±1.43	5.52±1.20	4.85±1.46
P3	Hallucinatory behavior	4.44±1.59	4.52±1.68	4.57±1.59	4.55±1.66
P4	Excitement	2.5±1.6	2.87±1.73	3.04±1.59	2.75±1.68
P5	Grandiosity	2.16±1.73	2.04±1.84	1.85±1.23	2.15±2.08
P6	Suspiciousness/persecution	4.20±1.52	4±1.711	4.42±1.77	3.85±1.72
P7	Hostility	2.65±1.86	2.71±1.59	3.23±1.92	2.65±1.63
G8	Uncooperativeness	3.96±1.63	2.54±1.54	3.14±2.00	2.6±1.69
G9	Unusual thought content	3.96±1.63	4.07±1.8	4.38±1.56	4.05±1.73
G14	Poor impulse control	2.40±1.75	2.73±1.60	3.38±1.96	2.55±1.63
G16	Active social avoidance	2.53±1.69	3±1.68	3.85±1.98	2.3±1.49

Table 44: External validity of classes of patients at the acute state by the four-class RLCA model with divisive hierarchical clustering method

	Class 1 vs. Class4		Class 2 vs. Class4		Class 3 vs. Class 4	
Variable[b]	OR[c]	CI[c]	OR	CI	OR	CI
Female gender	**3.681**	(1.229,11.022)	2.682	(0.795,9.047)	1.221	(0.300,4.974)
Age (1 year)	1.015	(0.928,1.111)	1.046	(0.945,1.157)	0.901	(0.783,1.036)
Age of onset (1 year)	0.940	(0.771,1.146)	0.816	(0.656,1.015)	0.978	(0.756,1.266)
Years of education (1 year)	1.197	(0.343,4.175)	1.565	(0.405,6.042)	0.869	(0.172,4.389)
Having occupation	1.025	(0.9249,1.136)	1.010	(0.902,1.121)	1.102	(0.948,1.282)
Brain injury (no)						
Slight and obvious[d]	1.341	(0.240,7.481)	0.984	(0.075,12.821)	0.738	(0.052,10.287)
Abnormal behavior (no)						
Slight	0.975	(0.245,3.879)	1.563	(0.359,6.790)	0.758	(0.128,4.479)
Obvious	0.348	(0.007,1.647)	**0.132**	(0.001,0.921)	0.1	(0.001,1.406)
Psychological problem (no)						
Slight	0.996	(0.226,4.372)	1.972	(0.395,9.826)	0.835	(0.115,6.041)
Obvious	1.897	(0.397,9.050)	2.501	(0.437,14.306)	2.476	(0.344,17.785)
CPT (1 unit)						
Undegraded d'	1.140	(0.841,1.543)	1.186	(0.8404,1.675)	1.473	(0.992,2.189)

[a] Class 4 was used as the reference class in calculating odds ratios.
[b] Parentheses identify the unit of increase or the reference group for which the odds ratio was calculated.
[c] OR, odds ratio; CI, 95% confidence interval of OR.
* *p*-value < 0.05

Table 45: Composition of classes of patients at the subsided state by the three-class RLCA model with divisive hierarchical clustering method

Factor	Subtype	Class 1	Class 2	Class 3
F1	Symptom	**15.35±7.27**	**13.19±5.05**	**13.55±6.11**
N1	Blunted affect	2.65±1.29	2.28±1.13	2.31±1.25
N2	Emotional withdrawal	2.40±1.40	2.04±1.05	2.05±1.18
N3	Poor rapport	2.12±1.33	1.69±1.02	1.71±1.01
N4	Passive/apathetic social withdrawal	2.76±1.43	2.28±1.04	2.52±1.24
N6	Lack of spontaneity/flow of conversation	2.35±1.53	2±1.14	2±1.18
G7	Motor retardation	1.70±0.94	1.47±0.74	1.63±1.07
G10	Disorientation	1.33±0.80	1.40±0.73	1.61±0.77
F2		**22.77±9.14**	**21.07±8.47**	**20.61±9.33**
P1	Delusions	2.74±1.65	2.85±1.55	0.42±1.57
P2	Conceptual disorganization	2.28±1.38	2.02±1.22	2.10±1.41
P3	Hallucinatory behavior	2.71±1.74	2.31±1.55	2.18±1.48
P6	Suspiciousness/persecution	1.04±1.27	2.09±1.44	2.07±1.30
N5	Difficulty in abstract thinking	3.23±1.37	2.88±1.53	2.65±1.51
N7	Stereotyped thinking	2.38±1.38	1.83±1.16	2.07±1.47
G9	Unusual thought content	2.44±1.77	2.54±1.65	2.47±1.68
G12	Lack of judgment and insight	3.12±1.31	2.78±1.31	2.76±1.19
G16	Active social avoidance	1.79±1.04	1.73±0.96	1.84±1.07
F3		**14.07±5.08**	**13±5.08**	**13.87±5.97**
G1	Somatic concern	1.82±1.05	1.88±1.04	2.55±1.51
G2	Anxiety	2.08±1.18	1.83±1.14	1.86±1.01
G4	Tension	1.60±0.90	1.61±0.93	1.78±1.25
G5	Mannerisms and posturing	1.33±0.90	1.26±0.79	1.36±0.78
G8	Uncooperativeness	1.50±0.97	1.40±0.91	1.42±0.82
G11	Poor attention	1.79±0.90	1.66±0.92	1.65±0.99
G13	Disturbance of volition	2.12±1.20	1.71±0.97	1.73±1.05
G15	Preoccupation	1.80±1.11	1.61±0.93	1.47±1.00
F4		**9.13±3.44**	**9.28±2.97**	**8.60±3.30**
P4	Excitement	1.63±0.98	1.40±0.82	1.31±0.66
P5	Grandiosity	1.48±1.20	1.45±0.94	1.57±1.13
P7	Hostility	1.45±1.07	1.42±0.77	1.42±0.85
G3	Guilt feelings	1.30±0.71	1.81±1.11	1.44±1.10
G6	Depression	1.67±0.96	1.57±0.99	1.5±0.98
G14	Poor impulse control	1.58±0.99	1.61±0.93	1.34±0.70

Table 46: External validity of classes of patients at the subsided state by the three-class RLCA model with divisive hierarchical clustering method

Variable[b]	Class 1 vs. Class3		Class 2 vs. Class3	
	OR[c]	CI[c]	OR	CI
Female gender	0.954	(0.419,2.169)	0.472	(0.175, 1.273)
Age (1 year)	1.017	(0.956,1.081)	1.031	(0.960,1.107)
Age of onset (1 year)	**0.831**	(0.708,0.975)	0.908	(0.758,1.088)
Years of education (1 year)	0.779	(0.315,1.922)	0.521	(0.174, 1.559)
Having occupation	0.985	(0.909,1.068)	0.995	(0.191, 1.090)
Brain injury (no)				
Slight and obvious[d]	1.680	(0.580,4.865)	2.340	(0.714,7.668)
Abnormal behavior (no)				
Slight	0.837	(0.306,2.292)	0.635	(0.191, 2.108)
Obvious	**0.216**	(0.065,0.717)	0.271	(0.065, 1.129)
Psychological problem (no)				
Slight	0.910	(0.293,2.816)	**5.305**	(1.570,17.926)
Obvious	2.272	(0.736,7.009)	1.739	(0.422,7.165)
CPT (1 unit)				
Undegraded d'	1.227	(0.921,1.634)	1.286	(0.912, 1.815)

[a] Class 3 was used as the reference class in calculating odds ratios.

[b] Parentheses identify the unit of increase or the reference group for which the odds ratio was calculated.

[c] OR, odds ratio; CI, 95% confidence interval of OR.

* *p*-value < 0.05

Table 47: Composition of classes of patients at the acute state by the four-class RLCA model with k-means clustering method

Factor	Subtype	Class 1	Class 2	Class 3	Class 4
F1	Symptom	**7.88±2.15**	**8.59±2.77**	**9.81±2.13**	**8.17±2.50**
N5	Difficulty in abstract thinking	3.32±1.55	3.63±1.97	4.62±1.44	3.52±1.78
G12	Lack of judgment and insight	4.55±1.31	4.95±1.46	5.18±1.21	4.64±1.24
F2		**20.19±6.46**	**14.10±6.64**	**29.37±8.01**	**16.05±6.97**
N1	Blunted affect	3.09±1.06	2.58±1.58	4.46±1.40	2.5±1.36
N2	Emotional withdrawal	3.55±1.48	2.39±1.49	4.79±1.16	2.80±1.53
N3	Poor rapport	2.69±1.40	1.97±1.35	4.44±1.77	2.14±1.31
N4	Passive/apathetic social withdrawal	3.25±1.36	2.48±1.55	4.55±1.66	2.64±1.52
N6	Lack of spontaneity/flow of conversation	2.69±1.40	1.82±1.24	4.37±1.77	2.16±1.39
G7	Motor retardation	2.23±1.13	1.29±0.71	3.20±1.61	1.66±1.07
G13	Disturbance of volition	2.65±1.27	1.53±1.09	3.53±1.40	2.11±1.17
F3		**15.49±5.47**	**10.49±5.31**	**23.09±5.07**	**13.57±5.64**
P2	Conceptual disorganization	3.34±1.36	2.24±1.65	4.46±1.24	3.11±1.67
N7	Stereotyped thinking	2.91±1.39	1.53±1.16	4.02±1.42	2.38±1.49
G5	Mannerisms and posturing	1.86±1.55	1.19±0.64	3.04±1.67	1.61±1.41
G10	Disorientation	1.58±1.01	1.46±0.95	2.9±1.46	1.57±1.03
G11	Poor attention	2.44±1.20	2.04±1.41	4.16±1.27	2.02±1.23
G15	Preoccupation	3.34±1.37	2±1.44	4.48±1.31	2.85±1.84
F4		**12.98±5.47**	**8.85±3.40**	**12.33±4.43**	**10±3.66**
G1	Somatic concern	3.11±1.62	1.97±1.19	2.30±1.56	2.45±1.53
G2	Anxiety	3±1.43	2.24±1.13	3.25±1.61	2.30±1.21
G3	Guilt feelings	1.91±1.29	1.31±0.98	1.51±0.79	1.33±0.84
G4	Tension	2.51±1.42	1.51±0.84	2.95±1.37	1.83±1.21
G6	Depression	2.44±1.40	1.80±1.05	2.30±1.24	2.07±1.21
F5		**32.30±9.94**	**30±8.99**	**41.86±11.06**	**29.88±8.47**
P1	Delusions	5.23±1.29	4.90±1.44	5.16±1.42	4.97±1.45
P3	Hallucinatory behavior	4.27±1.76	4.43±1.73	4.81±1.31	4.42±1.62
P4	Excitement	2.74±1.57	2.12±1.41	3.41±1.84	2.45±1.45
P5	Grandiosity	1.93±1.68	2.29±2.02	2.16±1.74	2±1.53
P6	Suspiciousness/persecution	3.95±1.47	4.29±1.53	4.48±1.48	3.83±1.92
P7	Hostility	2.48±1.57	2.21±1.40	3.95±1.82	2.26±1.71
G8	Uncooperativeness	2.39±1.54	1.95±1.30	4.02±1.76	2.09±1.44
G9	Unusual thought content	4.11±1.59	3.85±1.96	4.51±1.43	3.71±1.59
G14	Poor impulse control	2.46±1.68	2.04±1.35	3.69±1.92	2.26±1.51
G16	Active social avoidance	2.69±1.41	1.87±1.28	4.62±1.49	1.88±1.21

Table 48: External validity of classes of patients at the acute state by the four-class RLCA model with k-means clustering method

	Class 1 vs. Class4		Class 2 vs. Class4		Class 3 vs. Class 4	
Variable[b]	OR[c]	CI[c]	OR	CI	OR	CI
Female gender	**2.632**	(1.019,6.795)	**4.689**	(1.800,12.213)	1.989	(0.973,4.992)
Age (1 year)	1.014	(0.935,1.100)	0.985	(0.905,1.073)	1.045	(0.966,1.130)
Age of onset (1 year)	1.119	(0.946,1.322)	1.095	(0.925,1.295)	1.65	(0.906,1.251)
Years of education (1 year)	0.685	(0.247,1.900)	0.473	(0.925,1.295)	0.818	(0.308,2.171)
Having occupation	0.999	(0.911,1.095)	1.026	(0.933,1.129)	0.971	(0.890,1.060)
Brain injury (no)						
Slight and obvious[d]	0.779	(0.134,4.508)	0.479	(0.067,3.391)	0.506	(0.075,3.410)
Abnormal behavior (no)						
Slight	2.225	(0.700,7.260)	1.212	(0.353,4.166)	1.473	(0.450,4.818)
Obvious	0.594	(0.110,3.191)	0.747	(0.173,3.211)	0.595	(0.132,26.72)
Psychological problem (no)						
Slight	0.448	(0.108,1.853)	1.197	(0.327,4.373)	1.133	(0.312,4.047)
Obvious	0.608	(0.153,2.410)	1.231	(0.330,4.594)	0.907	(0.238,3.453)
CPT (1 unit)						
Undegraded d'	0.836	(0.632,1.106)	0.984	(0.745,1.298)	0.979	(0.746,1.285)

[a] Class 4 was used as the reference class in calculating odds ratios.
[b] Parentheses identify the unit of increase or the reference group for which the odds ratio was calculated.
[c] OR, odds ratio; CI, 95% confidence interval of OR.
* *p*-value < 0.05

Table 49: Composition of classes of patients at the subsided state by the three-class RLCA model with k-means clustering method

Factor	Subtype	Class 1	Class 2	Class 3
F1	Symptom	**18.16±6.92**	**12.88±5.6**	**10.87±3.87**
N1	Blunted affect	3.12±1.33	2.21±1.11	1.89±0.81
N2	Emotional withdrawal	2.88±1.32	2.05±1.20	1.53±0.77
N3	Poor rapport	2.52±1.34	1.72±1.04	1.29±0.68
N4	Passive/apathetic social withdrawal	3.25±1.20	2.41±1.39	1.87±0.94
N6	Lack of spontaneity/flow of conversation	2.74±1.52	1.92±1.32	1.7±0.85
G7	Motor retardation	1.96±1.12	1.43±0.75	1.38±0.64
G10	Disorientation	1.65±0.98	1.11±0.43	1.19±0.61
F2		**29±7.68**	**19.35±6.94**	**14.85±4.88**
P1	Delusions	3.63±1.66	2.45±0.33	1.70±1.01
P2	Conceptual disorganization	3.01±1.41	1.86±1.13	1.38±0.67
P3	Hallucinatory behavior	3.39±1.68	2.31±1.40	1.44±1.09
P6	Suspiciousness/persecution	2.90±1.39	1.68±0.94	1.36±0.89
N5	Difficulty in abstract thinking	3.74±1.33	2.58±1.38	2.46±1.29
N7	Stereotyped thinking	2.82±1.42	1.92±1.29	1.55±0.95
G9	Unusual thought content	3.55±1.81	2.13±1.41	1.4±0.85
G12	Lack of judgment and insight	3.63±1.08	2.72±1.28	2.27±1.13
G16	Active social avoidance	2.28±1.08	1.66±0.97	1.25±0.64
F3		**18.23±5.19**	**11.66±3.12**	**9.97±1.87**
G1	Somatic concern	2.53±1.30	1.84±1.17	1.48±0.77
G2	Anxiety	2.61±1.19	1.58±0.87	1.51±0.88
G4	Tension	2.14±1.09	1.31±0.88	1.36±0.70
G5	Mannerisms and posturing	1.66±1.10	1.19±0.72	1±0
G8	Uncooperativeness	1.87±1.12	1.31±0.81	1.06±0.32
G11	Poor attention	2.43±0.98	1.47±0.70	1.17±0.48
G13	Disturbance of volition	2.60±1.17	1.58±0.92	1.38±0.38
G15	Preoccupation	2.44±1.17	1.35±0.71	1±0
F4		**11.12±3.76**	**8.37±2.14**	**7±1.64**
P4	Excitement	1.92±1.00	1.37±0.84	1.06±0.32
P5	Grandiosity	1.87±1.49	1.41±0.82	1.08±0.45
P7	Hostility	1.95±1.22	1.15±0.61	1.06±0.32
G3	Guilt feelings	1.57±1.01	1.50±0.98	1.29±0.80
G6	Depression	1.85±1.02	1.50±0.98	1.38±0.79
G14	Poor impulse control	1.95±1.08	1.41±0.82	1.10±0.42

Table 50: External validity of classes of patients at the subsided state by the three-class RLCA model with k-means clustering method

	Class 1 vs. Class3		Class 2 vs. Class3	
Variable	OR	CI	OR	CI
Female gender	1.465	(0.647,3.316)	1.666	(0.742,3.742)
Age (1 year)	0.993	(0.935,1.055)	0.956	(0.899,1.016)
Age of onset (1 year)	**0.818**	(0.696,0.961)	0.987	(0.847,1.149)
Years of education (1 year)	1.527	(0.602,3.873)	1.593	(0.634,4.004)
Having occupation	0.981	(0.909,1.059)	0.996	(0.919,1.080)
Brain injury (no)				
Slight and obvious	0.673	(0.258,1.759)	0.674	(0.267,1.702)
Abnormal behavior (no)				
Slight	1.188	(0.450,3.139)	1.034	(0.383,2.789)
Obvious	0.534	(0.155,1.838)	0.990	(0.311,3.145)
Psychological problem (no)				
Slight	0.739	(0.263,2.078)	0.712	(0.248,2.044)
Obvious	0.828	(0.274,2.500)	1.246	(0.431,3.596)
CPT (1 unit)				
Undegraded d'	1.320	(0.986,1.767)	1.100	(0.830,1.459)

[a] Class 3 was used as the reference class in calculating odds ratios.

[b] Parentheses identify the unit of increase or the reference group for which the odds ratio was calculated.

[c] OR, odds ratio; CI, 95% confidence interval of OR.

* p-value < 0.05

Table 51: External validity of classes of breast cancer patients by the two-class RLCA model with divisive hierarchical clustering method

Variable[b]	Class 1 vs. Class2 OR[c]	CI[c]
Age (1 year)	0.95	(0.873,1.034)
Metastases more than 5 year	14.154*	(4.076,49.148)

[a] Class 2 was used as the reference class in calculating odds ratios.
[b] Parentheses identify the unit of increase or the reference group for which the odds ratio was calculated.
[c] OR, odds ratio; CI, 95% confidence interval of OR.
* *p*-value < 0.05

Table 52: Predictions of class membership of 19 tumours by divisive hierarchical clustering method

Individual	Prediction class membership*	Posterior Probability class 1	Posterior Probability class 2	True class membership
1	2	0.154	0.846	2
2	2	0.002	0.998	2
3	**1**	**0.712**	**0.288**	2
4	1	0.999	0.001	1
5	1	0.962	0.038	1
6	1	0.999	0.001	1
7	1	0.999	0.001	1
8	1	0.998	0.002	1
9	**2**	**0.002**	**0.998**	1
10	1	0.813	0.187	1
11	2	0.001	0.999	2
12	2	0.001	0.999	2
13	2	0.001	0.999	2
14	**1**	**0.999**	**0.001**	2
15	2	0.001	0.999	2
16	2	0.001	0.999	2
17	2	0.007	0.993	2
18	2	0.001	0.999	2
19	2	0.001	0.999	2

*Values in bold are misclassification

Table 53: External validity of classes of breast cancer patients by the two-class RLCA model with k-means clustering method

Variable[b]	Class 1 vs. Class2 OR[c]	CI[c]
Age (1 year)	0.95	(0.873,1.034)
Metastases more than 5 year	14.154*	(4.076,49.148)

[a] Class 2 was used as the reference class in calculating odds ratios.
[b] Parentheses identify the unit of increase or the reference group for which the odds ratio was calculated.
[c] OR, odds ratio; CI, 95% confidence interval of OR.
* *p*-value < 0.05

Table 54: Predictions of class membership of 19 tumours by k-means clustering method

Individual	Prediction class membership*	Posterior Probability class 1	Posterior Probability class 2	True class membership
1	2	0.154	0.846	2
2	2	0.002	0.998	2
3	**1**	**0.712**	**0.288**	2
4	1	0.999	0.001	1
5	1	0.962	0.038	1
6	1	0.999	0.001	1
7	1	0.999	0.001	1
8	1	0.998	0.002	1
9	**2**	**0.002**	**0.998**	1
10	1	0.813	0.187	1
11	2	0.001	0.999	2
12	2	0.001	0.999	2
13	2	0.001	0.999	2
14	**1**	**0.999**	**0.001**	2
15	2	0.001	0.999	2
16	2	0.001	0.999	2
17	2	0.007	0.993	2
18	2	0.001	0.999	2
19	2	0.001	0.999	2

*Values in bold are misclassification

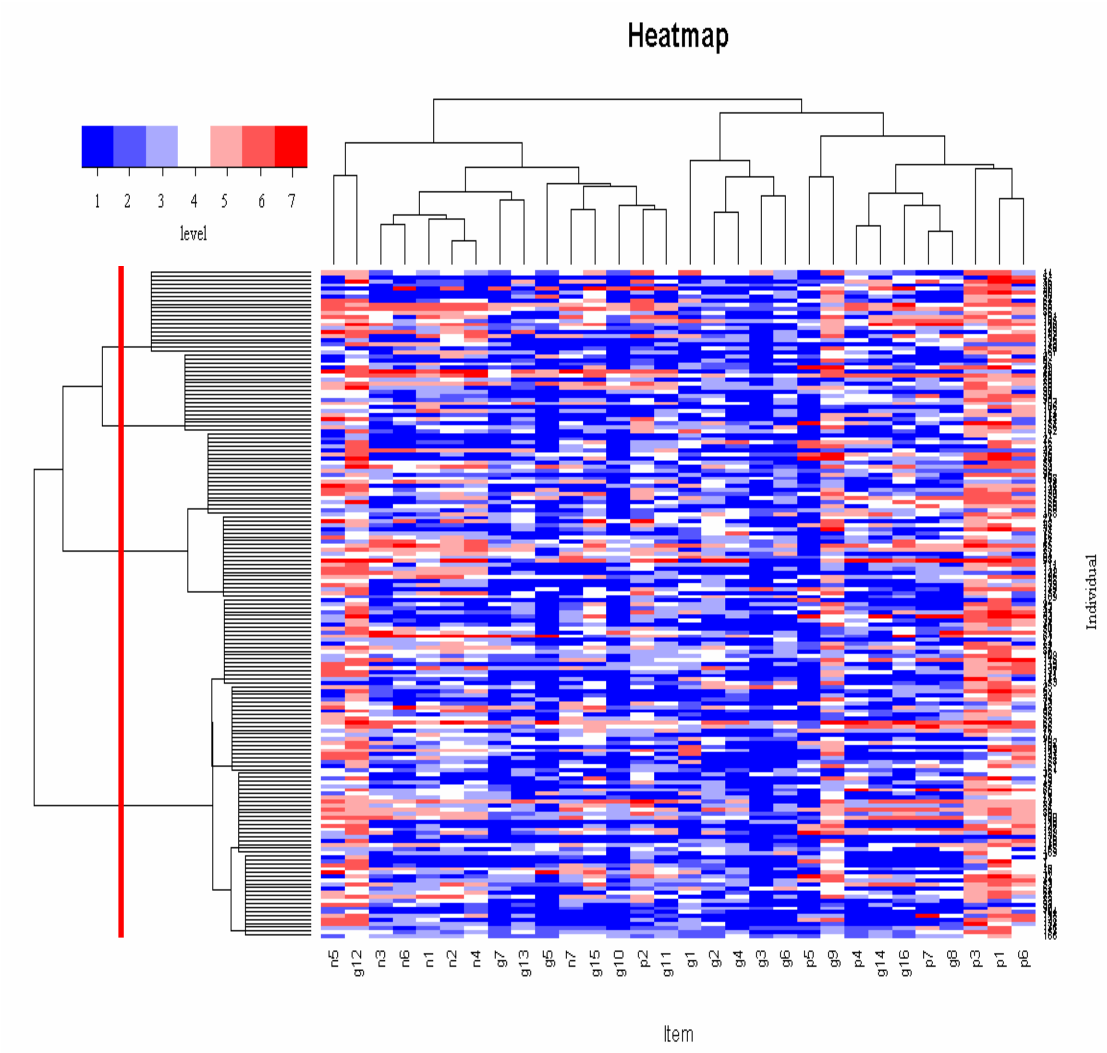

Figure 3: Heatmap for schizophrenia patients at the acute state

Two dimensional presentation of the level of item for schizophrenia patients at the acute state. There were 30 items across the group. Each row represents a individual and each column a single item. As shown in the color bar, blue indicates level 1, red level 7

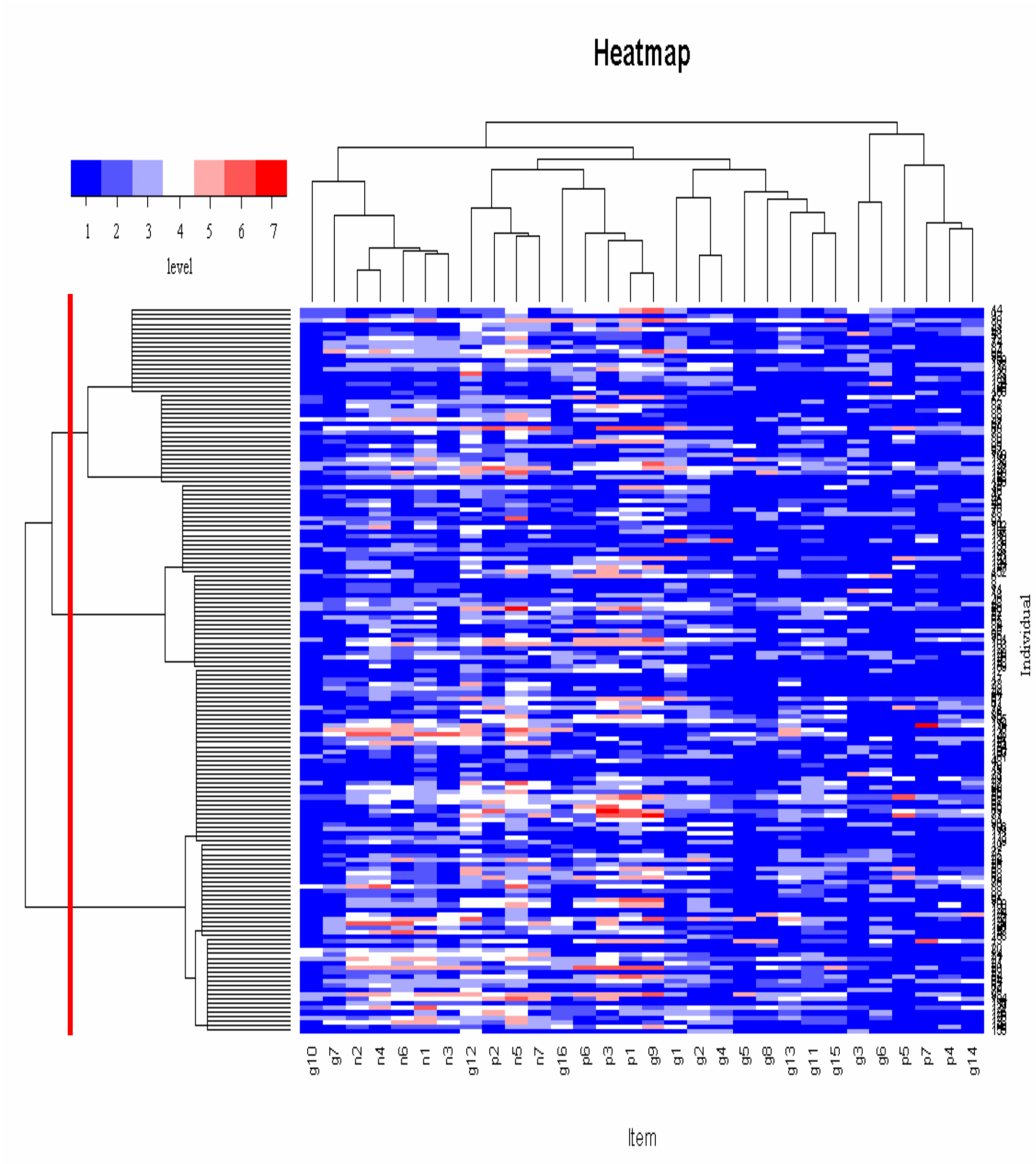

Figure 4: Heatmap for schizophrenia patients at the subsided state

Two dimensional presentation of the level of item for schizophrenia patients at the subsided state. There were 30 items across the group. Each row represents a individual and each column a single item. As shown in the color bar, blue indicates level 1, red level 7

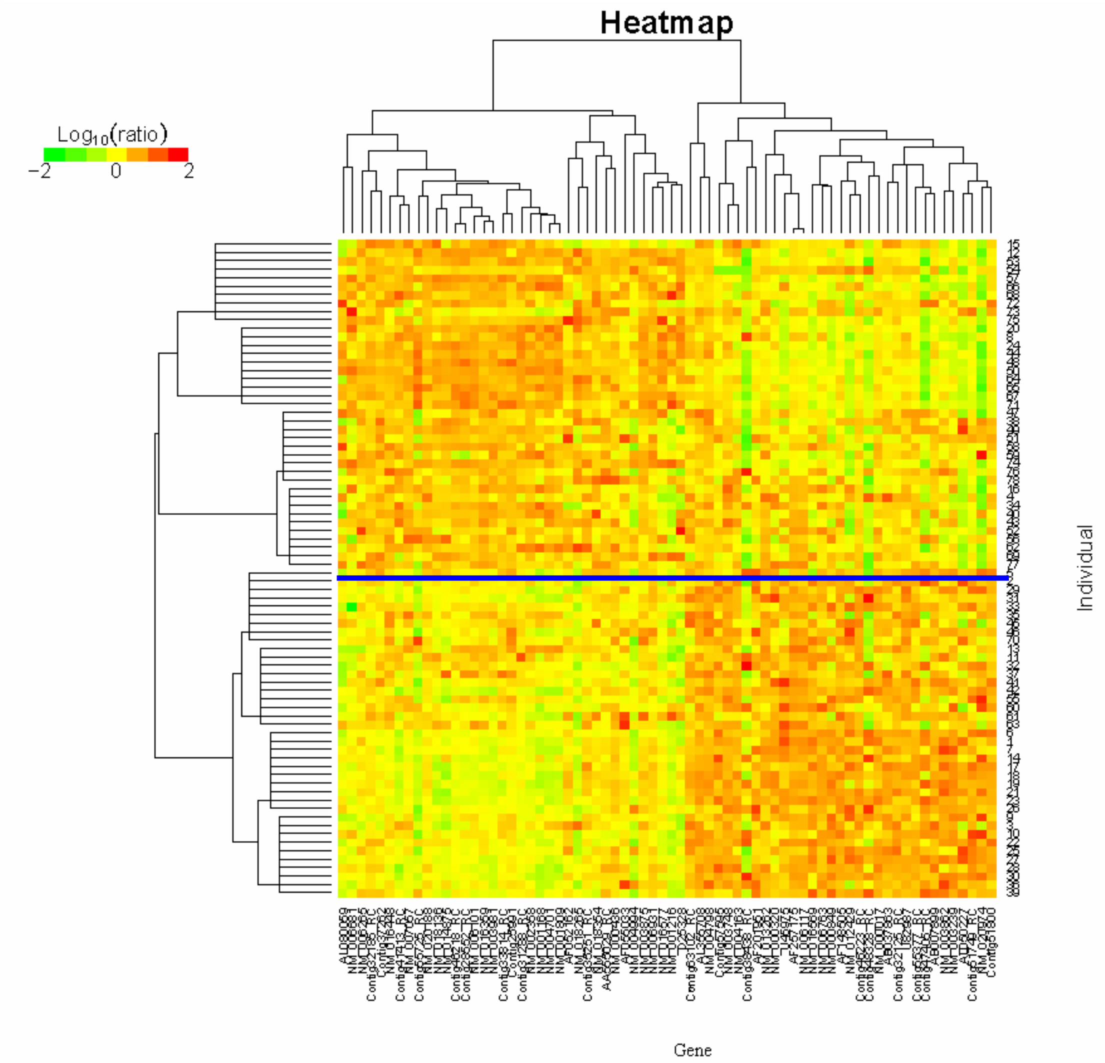

Figure 5: Heatmap for breast tumours

Two dimensional presentation of transcript ratios for 78 breast tumours. There were 70 significant genes across the group. Each row represents a individual and each column a single gene. As shown in the color bar, red indicates upregulation, green downregulation, yellow no change