Open Access

An Application of Latent Class Analysis in the Measurement of Falling Among a Community Elderly Population

Patrick C. Hardigan*

Statistical Consulting Center, Nova Southeastern University, 3200 South University Drive, Ft. Lauderdale, FL 33328, USA

Abstract: *Purpose*: Latent Class Analysis (LCA) is a statistical method for finding subtypes of related cases (latent classes) from multivariate categorical data. LCA is well suited to many health applications where one wishes to identify disease subtypes or diagnostic subcategories. In this paper we demonstrate the utility of LCA for the prediction of falls among community dwelling elderly. Falls among the elderly are a major public health concern. Therefore the possibility of a modeling technique which could better estimate fall probability is both timely and needed.

Methods: A three-step modeling approach was employed. First, we looked for the optimal number of latent classes for the seven binary indicators: (1) arthritis, (2) high blood pressure, (3) diabetes, (4) heart disease, (5) foot disorders, (6) Parkinson's disease, and (7) stroke. Second, we modeled two covariates (age and number of medications) on the latent class. Third, we modeled the appropriate latent class structure, with the covariates, on the distal outcome (fall/no fall). The default estimator is maximum likelihood with robust standard errors. The Pearson chi-square, likelihood ratio chi-square, BIC, Lo-Mendell-Rubin Adjusted Likelihood Ratio test and the bootstrap likelihood ratio test are used for model comparisons.

Results: A review of the model fit indices with covariates shows that a five-class solution was preferred. The predictive probability for latent classes ranged from 60% to 72%. Persons in classes one, two and five possessed the greatest probability of falling.

Conclusions: In conclusion we found the LCA method effective for finding relevant subgroups with a heterogenous at-risk population for falling. This study demonstrated that LCA offers researchers a valuable tool to model medical data.

Keywords: Latent class analysis, elderly falling.

INTRODUCTION

Latent Class Analysis (LCA) is a statistical method for finding subtypes of related cases (latent classes) from multivariate categorical data.[1] The most common use of LCA is to discover case subtypes (or confirm hypothesized subtypes) based on multivariate categorical data [1-4]. LCA is well suited to many health applications where one wishes to identify disease subtypes or diagnostic subcategories [1-4]. LCA models do not rely on traditional modeling assumptions (normal distribution, linear relationship, homogeneity) and are therefore less subject to biases associated with data not conforming to model assumptions [1-4]. In this paper we demonstrate the utility of LCA for the prediction of falls among community dwelling elderly.

Falls among the elderly are a major public health concern. Research on falls and fall-related behavior among the elderly has found that falls are the leading cause of injury deaths among individuals who are over 65 years of age [5-11]. Research has shown that sixty percent of fall-related deaths occur among individuals who are 75 years of age or older [5-11]. Demography research estimates that by 2030, the population of individuals who are 65 years of age or older will double and by 2050 the population of individuals who are 85 years of age or older will quadruple [5-11].

Predicting elderly falling can be complex and often involves heterogeneous markers. Therefore the identification of more homogeneous subgroups of individuals and the refinement of the measurement criteria are typically interrelated research goals. Appropriate new statistical applications, such as latent class analysis, have become available for researchers to model the complex heterogenous measurements.

Latent class models are used to cluster participants. This type of model is adequate if the sample consists of different subtypes and it is not known before-hand which participant belongs to which of the subtypes [2]. The latent categorical variable is used to model heterogeneity. In the classic form of the latent class model, observed variables within each latent class are assumed to be independent, and no specific structure for the covariances of observed variables is specified [2].

LCA is one of the most widely used latent structure models for categorical data [12]. LCA differs from more well-known methods such as K-means clustering which apply arbitrary distance metrics to group individuals based on their

*Address correspondence to this author at the Statistical Consulting Center, Nova Southeastern University, 3200 South University Drive, Ft. Lauderdale, FL 33328, USA; Tel: 954-262-1524; Fax: 954-262-2252;
E-mail: patrick@nova.edu

similarity [13-15]. LCA derives clusters based on conditional independence assumptions applied to multivariate categorical data distributed as binomial or multinomial variables [16, 17]. Using statistical distributions rather than distance metrics to define clusters helps in evaluating whether a model with a particular number of clusters is able to fit the data, since tests can be performed to observed (ni) *vs* model expected values (mi), using exact methods as recommended [18, 19]. This comparison gives rise to a χ^2 test of global model fit, in which significant values index lack of fit [16]. Here lack of fit means deviation of (model) predicted (m) frequencies from observed frequencies (n) [16].

Latent class analysis assumes that each observation is a member of one and only one T latent (unobservable) and that the indicator (manifest) variables are mutually independent of each other [20]. The models are expressed in probabilities of belonging to each latent class. In this example, we have seven manifest variables which can be expressed as:

$$\pi_{ijklmnot} = \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X} \pi_{mt}^{E|X} \pi_{nt}^{F|X} \pi_{ot}^{G|X}$$

where π_t^X denotes the probability of being in a latent class (t=1,2,...,T) of latent variable X; $\pi_{it}^{A|X}$ denotes the conditional probability of obtaining the *i*th response from item A, from members of class t, i = 1,2,...,I; and $\pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X} \pi_{mt}^{E|X} \pi_{nt}^{F|X} \pi_{ot}^{G|X}$, j=1,2,...,j k=1,2,...,k l= 1,2,...,l m=1,2,...,m n=1,2,...,n O= 1,2,...,O are the corresponding conditional probabilities for items B,C,D,E,F, and G respectively.

We are testing the hypothesis that a two-class distal outcome (fall/no fall) can explain the relationship among the biomedical and demographic variables. Proper analysis of this data requires the understanding of two interdependent outcomes [2]. First, the binary outcome is whether or not the event occurred (fall or no fall) and second what covariates increase or decrease the likelihood of this occurrence. The four specific aims are to identify items that indicate classes, estimate class probabilities, relate the class probabilities to covariates, and to predict a distal outcome (fall/no-fall) based on class membership. We model this process through the application of latent class analysis (Fig. **1**).

METHODS

A data set consisting of 3293 elderly people was analyzed. Using correspondance analysis we analyzed the data set to identify potential manifest variables. Through this process the seven medical markers were identified. It should be noted that this data set was not designed for a latent class analysis, therefore additional medical variables which may predict falling were not available for analysis. For this study an elderly person is someone aged 65 years or older. Seventy-four percent of the subjects had not fallen while 26 percent had fallen in the last 30 days. Descriptive statistics for the data set are presented in Table **1**.

Table 1. Descriptive Statistics

		No Fall	Fall
Age	Mean ± SD	77.47 ± 6.91	77.98 ± 7.41
Medications	Mean ± SD	2.30 ± 5.57	5.10 ± 10.10
Gender	Male	27%	22%
	Female	73%	78%

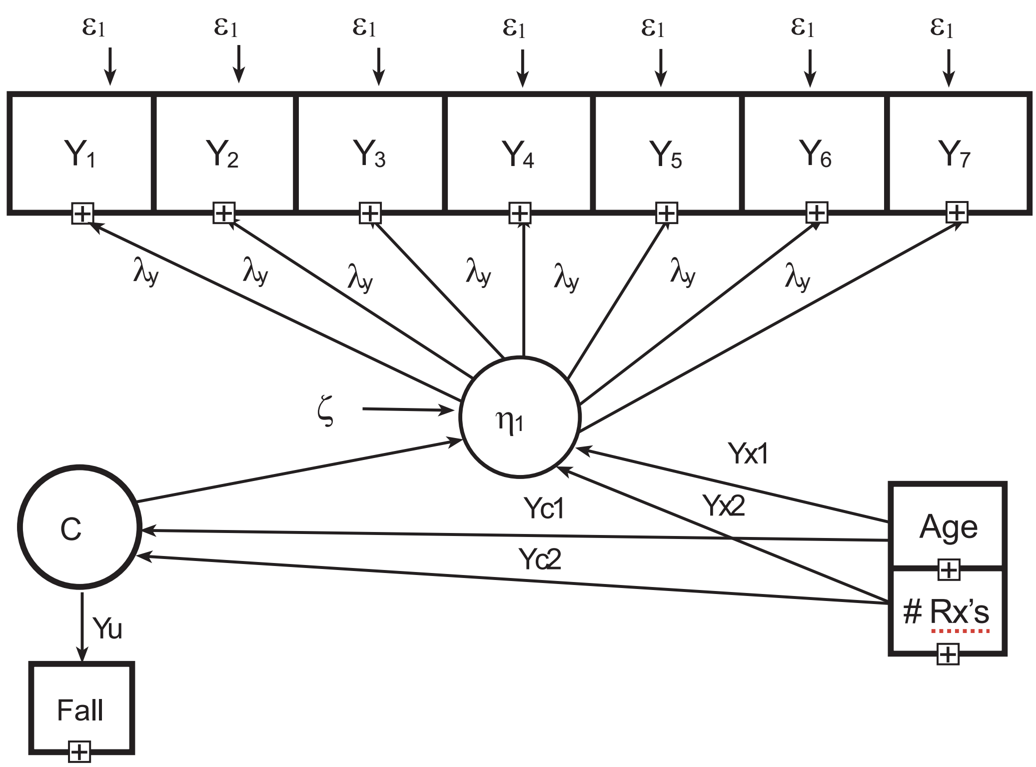

Fig. (1). Proposed model.

MEASUREMENTS

Distal Outcome

Falling is defined as "an event which results in the person coming to rest inadvertently on the ground or other lower level, and other than as a consequence of sustaining a violent blow."

- Binary Component
    - Coded as 1 if any and
    - 0 if none

Biomedical indicators, measured on a categorical level "yes/no," were identified by medical personnel.

- Arthritis
- High Blood Pressure
- Diabetes
- Heart Disease
- Foot Disorders
- Parkinson's Disease
- Stroke

Covariates

- Age
- Number of prescription medications

A three-step modeling approach was employed. First, we looked for the optimal number of latent classes for the seven binary indicators: (1) arthritis, (2) high blood pressure, (3) diabetes, (4) heart disease, (5) foot disorders, (6) Parkinson's disease, and (7) stroke. Second, we modeled two covariates (age and number of medications) on the latent class. Third, we modeled the appropriate latent class structure, with the covariates, on the distal outcome (fall/no fall). The default estimator is maximum likelihood with robust standard errors. The Pearson chi-square, likelihood ratio chi-square, BIC, Lo-Mendell-Rubin Adjusted Likelihood Ratio test and the bootstrap likelihood ratio test are used for model comparisons.

RESULTS

A review of the model fit indices with no covariates, shows that a five-class solution was preferred. The five-class solution provided a lower Bayesian Information Criteria--BIC (lower is better), much smaller chi-square values, and as indicated by the procedures (Lo-Mendell-Rubin likelihood ratio test--LMR and bootstrap likelihood ratio test--BLRT) the five-class solution was preferred (non-significant p-values). We then modeled the five-class solution, with the covariates age and number of prescription medications, on the distal outcome--fall or no fall. Age was shown to be non-significant but we left it in the model as age has been shown to significantly influence the probability of falling. The number of prescription medications possessed a significant impact on falling. The more prescription drugs an elderly person takes, the greater the probability that they will fall. Table **2** provides a comparison of fit indices for two-class, three-class, four-class, five-class, and six-class solutions. We interpret the five class structure, with covariates, as follows:

1. Class one is most likely to be affected by all medical conditions except for Parkinson's Disease (Fig. **2**). We define this as the poorest-health group. Nine percent of the sample is classified into latent class one (Table **3**). The predictive probability for latent class one is 72% (Table **4**). The misclassified elderly generally were placed into class two or class four. The odds ratio indicate that a person in class one is three times more likely to fall than a person in class three--the healthy group (Table **5**).
2. Class two is primarily affected by arthritis and to a lesser extent high blood pressure (Fig. **2**). We define this as the arthritis group. Eighteen percent of the sample were placed into latent class two (Table **3**). Of all the latent classes, class two possessed the poorest predictive validity (60%--Table **4**). Misclassified elderly were placed into either class three (healthy group) or four (high blood pressure group). The odds ratio indicate that a person in class two is about 2.5 times more likely to fall than a person in class three--the healthy group (Table **5**).
3. Class three is generally unaffected by all medical markers (Fig. **2**). We define this as the healthy group. The majority of elderly (31%--Table **3**) in our sample were classified into latent class three. Of all groups, class three possessed the best predictive validity (75%--Table **4**). Misclassified elderly were placed into either class two or four. We used this as our comparison group for the calculation of odds ratios for falling.
4. Class four is primarily affected by high blood pressure (HBP), but also arthritis and diabetes (Fig. **2**). We define this as the HBP group. Thirty-one percent of the sample fell into latent class four (Table **3**). Class four possessed adequate predictive validity (70%--Table **4**). Misclassified elderly were placed into either class two (arthritis) or three (healthy group). The odds ratio indicate that a person in class four is approximately 1.5 times more likely to fall than a person in class three--the healthy group (Table **5**).
5. Class five is primarily affected by Parkinson's Disease (PK); therefore, we define this as the Parkinson's group. Six percent of the sample fell into latent class five (Table **3**). Class five possessed average predictive validity (61%--Table **4**). The relatively poor classification rate is probably due to the small number of elderly who possessed Parkinson's Disease. Misclassified elderly were placed into either class four (HBP) or three (healthy group). The odds ratio indicates that a person in class five is 39 times more likely to fall than a person in class three--the healthy group (Table **5**). This strongly indicates that a person with Parkinson's Disease is very likely to fall.

DISCUSSION

This paper demonstrated the utility of LCA in the measurement of falling among community-dwelling elderly. The basic idea underlying LCA is a very simple one: some of the

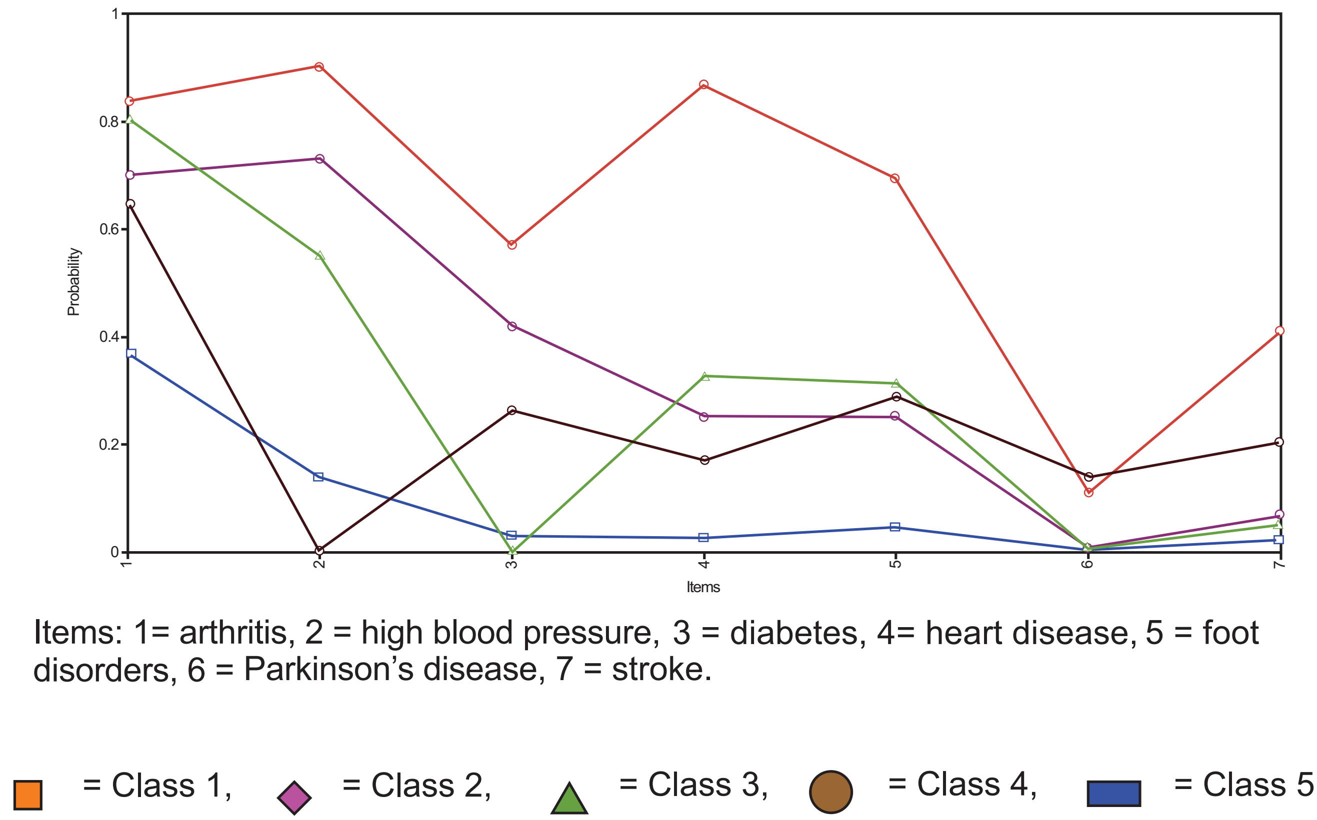

Fig. (2). Item Probabilities with covariates.

Table 2. Basic Latent Class Structure

	Two Class	Three Class	Four Class	Five Class	Six Class	5 Class With Covariates and Distal Outcome
Pearson χ^2	346	253	206	165	84	518
LR χ^2	346	204	153	109	84	411
χ^2 *df*	112	104	96	88	80	215
Loglikelihood	-10,591	-10,520	-10,494	-10,472	-10,460	-12,222
Number of Parameters	15	23	31	39	47	52
BIC	21,303	21,226	21,240	21,261	21,301	24,866
LMR (*p* value)	0.000	0.000	0.000	0.355	0.491	0.400
BLRT (*p* value)	0.000	0.000	0.000	0.359	0.495	0.410
Entropy	0.545	0.591	0.633	0.604	0.585	0.558

Table 3. Final Class Counts and Proportions Based on Posterior Probabilities

	Count	Proportion
1	312	9%
2	580	18%
3	1190	36%
4	1016	31%
5	195	6%
Total	**3293**	**100%**

Table 4. Most Likely Latent Class Membership

	Class 1	Class 2	Class 3	Class 4	Class 5
1	**72%**	7%	0%	19%	1%
2	5%	**60%**	11%	19%	5%
3	0%	11%	**75%**	10%	5%
4	7%	11%	9%	**70%**	3%
5	4%	8%	11%	16%	**61%**

Table 5. Odds Ratios

Class	Odds	Lower 95%	Upper 95%	P Value
1.00	3.32	2.49	4.15	0.00
2.00	2.63	2.03	3.40	0.00
4.00	1.63	1.32	2.03	0.00
5.00	39.32	21.64	71.45	0.00
Age	1.00	0.98	1.01	0.69
Number of Medications	1.01	1.00	1.09	0.00

Base or comparison group is class three or the "healthy" group.

parameters of a postulated statistical model differ across previously unrecognized subgroups [21]. These subgroups form the categories of a categorical latent variable. Given the variables the potential for confounding is great, but it is shown in past research that adjustment for a misclassified confounding variable is greatly improved by using the latent class class analysis [15].

The five-class solution was statistically sound and provided a relatively straightforward interpretable number of classes. The interpretation of a LCA relies on both the statistical indices and the practical interpretation of the classes. In our example the statistical indices strongly point towards a five factor model. The positive predictive values for both models (with and without a distal outcome) were acceptable. Furthermore, we were able to define each latent class, which provides researchers and practitioners practical implications of the analysis. The important idea, though, is to be able to identify the strengths and weaknesses of the model. For this discussion we will focus on the positive predictive values from the distal model--falling. Here it is interesting to note where the model was successful or unsuccessful in predicting class membership.

The model successfully predicted group membership for class three 75 percent of the time. The majority of elderly who should have been placed into class three were identified as persons in class four---high blood pressure group. A review of Fig. (**2**) demonstrates that the probability of item endorsement for class three and four generally parallels each other. The primary difference between the groups is the likelihood that an elderly patient possess a specific medical condition. Elderly community dwellers in class four possess a greater probability of having each medical condition than elderly in class three. Given that elderly in class four possess greater odds of falling, this would make rational sense.

The model had a difficult time placing persons into class two and five. Class two is the arthritis group and class five is the Parkinson's Disease group. The majority of misclassified elderly for the class two and five were placed into class four (19% and 16% respectively). This misclassification may be due to the fact that all three groups, classes two, four and five, possess high probabilities of having arthritis. Examining the odds ratios indicates that class two and four possess similar odds of falling, but class five has much greater odds. Class five's greater likelihood of falling is primarily due to Parkinson's Disease. Therefore, it is probable given additional biomedical conditions these three classes would better differentiate.

Class one possessed above average classification rate of 72 percent (Table **4**). A review of the item probability measures (Fig. **2**) shows that the elderly in class one possess the greatest likelihood of possessing all medical conditions except for Parkinson's Disease. Furthermore, elderly in class one had the second highest probability of falling. Similar to all other classes, misclassified elderly were placed into class four (19%). This pattern of misclassification appears to suggest that because many of the elderly possess the medical conditions, adding additional items or covariates may help to clarify the model. Nevertheless, we conclude that the five-class solution is a sound model and it provides us with good predictive validity for elderly falling.

CONCLUSION

In conclusion we found the LCA method effective for finding relevant subgroups with a heterogenous at-risk population for falling. This study demonstrated that LCA offers researchers a valuable tool to model medical data. Nevertheless, this study possesses limitations inherent to database research. Furthermore, the items were not designed for a LCA approach. A latent class study designed apriori may offer better solutions. We would also suggest that additional items (medical) be used which have demonstrated to impact falling among elderly community dwellers--such as eye disease.

REFERENCES

[1] Uebersax JS. LCA Frequently asked questions (FAQ). 4 March 2008. [Accessed Dec 1, 2008]. Available from http://ourworld.com puserve.com/homepages/ jsuebersax/faq.htm
[2] Hagenaars JA, McCutcheon AL, Eds. Applied latent class analysis. New York: Cambridge University Press 2002.
[3] Lazarsfeld PF, Henry NW. Latent structure analysis. Boston: Houghton 1968.
[4] McCutcheon AC. Latent class analysis. Beverly Hills: Sage 1987.
[5] Falls Prevention Act of 2003. §S. 1217.
[6] Cummings RG. Epidemiology of medication-related falls and fractures in the elderly. Drugs Aging 1998; 12(1): 43-53.
[7] Kannus Lord SR, Sherrington C, Menz H. Falls in older people: risk factors and strategies for prevention. Cambridge: Cambridge University Press 2001.
[8] Weir E, Culmer L. Fall prevention in the elderly population. CMAJ 2004; 171: 724.
[9] Tinetti ME. Preventing falls in elderly persons. N Engl J Med 2003; 348: 42-9.
[10] Kannus P, Sievnen H, Palvanen M, *et al.* Prevention of falls and consequent injuries in elderly people. Lancet 2005; 366: 1885-93.
[11] Kannus P, Palvanen M, Niemi S. Time trends in severe head injuries among elderly Finns. JAMA 2001; 286: 673-4.
[12] Croudace TJ, Jarvelin MR, Wadsworth ME, Jones PB. Developmental typology of trajectories to nighttime bladder control: epidemiologic application of longitudinal latent class analysis. Am J Epidemiol 2003; 157: 834-42.
[13] Everitt BS. The analysis of contingency tables. London: Chapman & Hall 1992.
[14] Everitt BS, Hand DJ. Finite mixture distributions. London: Chapman & Hall 1991.
[15] Magidson J, Vermunt JK. Latent class models for clustering: a comparison with *K*-means. Can J Mark Res 2002; 20: 37-44.
[16] Ploubidis GB, Abbott RA, Huppert FA, Kuh D, Wadsworth EJ, Croudace TJ. Improvements in social functioning reported by a birth cohort in mid-adult life: a person-centred analysis of GHQ-28 social dysfunction items using latent class analysis. Pers Individ Dif 2007; 42: 305-16.

[17] Clogg CC. In: Arminger G, Clogg CC, Sobel ME, Eds. Handbook of statistical modeling for the social and behavioral sciences. New York: Plenum Press 1995; pp. 311-359.
[18] Read TR, Cressie N. Goodness-of-fit statistics for discrete multivariate data. New York: Springer 1988.
[19] Langeheine R, Pannekoek J, van de Pol F. Bootstrapping goodness-of-fit measures in categorical data analysis. Sociol Methods Res 1996; 24: 492-516.
[20] Magidson J, Vermunt J. Latent class models. March 18, 2003. [Accessed Dec 1, 2008]. Available from http://www.statisticalinnovations.com/articles/sage 11.pdf
[21] Lewis-Beck MS, Bryman A, Liao TF, Eds. The Sage encyclopedia of social science research methods. New York: Sage Publications Inc. 2004.

Received: February 2, 2009 Revised: February 19, 2009 Accepted: February 24, 2009