Exploration of Groups Through Latent Structural Model

Bushra Shamshad* and Junaid Saghir Siddiqi

Department of Statistics, University of Karachi, Karachi, Pakistan

Abstract: In this paper Latent Class Analysis is applied on two different data sets. One of which is of elections of Karachi University Teacher Society (KUTS) in year 1993-1994. Members of two (Rightist and Mix) groups were competing for the post of President, Vice president, Secretary and Treasurer. The second data is about the study of parenting style on rearing children along with the factors self esteem and thoughts of suicide. From both the data set we will be able to come across the groups prevailing in our society and be able to assign conditional probability to individual, to which group they belong.

Keywords: Latent variable, Likelihood ratio statistic, EM algorithm, Information criterions (AIC and BIC).

1. INTRODUCTION

When dealing with large data such as surveys in which there are a bulk of information, it is difficult to handle it. Moreover simple descriptive statistic(s) is/are not enough to understand the information completely. There can be many hidden factors, which can not be measured directly. These factors can be explored with the help of many other measurable quantities, which can be directly observed.

For the problems of handling a great bulk of information, there are many data reduction techniques in multivariate analysis, for both continuous and categorical data. These dimension reduction techniques reduce the dimension of the large data without loosing much information.

Factor Analysis (FA) and Latent Class Analysis (LCA) are the data reduction techniques. FA is used for continuous variables [1], and LCA also known as Latent structure Model is used for dichotomous or multi-category variables [2]. The concept of LCA is similar to that of partial correlation. In case of partial correlation a third variable, influencing the relationship between another two variables, if kept constant, gives true relationship.

Similarly, a number of variables which can not be directly measured (latent variables) or for which one knows that the correct response could not be obtained by asking question directly [3] are introduced in the set of manifest variables (which are directly observable), in such a way that after introducing the latent variable the set of variables become independent (when applying Chi-square test of independence).

In Latent structure model [4-7], one of the main task is to identify the number of latent classes, which could be obtained by applying the concept of local independence. Therefore, one needs to start from the minimum number of classes and check the model adequacy. The process of adding latent classes continue until the assumption of conditional independence achieved. See also [8-10] for further study. Procedure for estimating parameters requires iterative solution in order to maximize the likelihood function. The most common method is EM Algorithm [11].

2. APPLICATION

Latent structure analysis is applied to the following data sets through poLCA package [12-14], in R statistical computing environment. Likelihood ratio statistics (G^2) and Chi-square goodness of fit (X^2) are calculated to asses the goodness of fit of the models [15]. G^2 and X^2 decreases with the increase of the number of classes to a latent class model and hence each additional class increase the fit of the model.

Parsimony measures Akaike information criteria (AIC), [16], and Bayesian information criteria (BIC), [17], are also calculated to check the model adequacy. It is important to keep in mind that if the number of parameter estimated exceeds the total number of observation the model will be unidentifiable. A model with minimum AIC and BIC along with minimum plausible estimated parameter is then selected.

2.1. KUTS Election Data

The data is taken from elections of Karachi University Teacher Society in year 1993-1994. At that time, there were 434 teachers, who vote in favor of one of the two groups. We name these groups as “Rightist”

*Address corresponding to this author at the Department of Statistics, University of Karachi, Karachi, Pakistan; Tel: 99261300-7, ext. 2298; E-mails: bushra_shamshad@yahoo.com, or jssdr123@yahoo.com

ISSN: 1814-8085 / E-ISSN: 1927-5129/12

and "Mix" group. From each group, individuals compete for President, Vice president, Secretary and Treasurer.

Originally, members elected from the Rightist group were for the post of President, Secretary and Treasurer (see Table **1**). Vice president is the only one elected from mix group. In this paper we consider positive response for voting in favor of Rightist group.

Table 1: Original Results of 1993-1994 KUTS Election

	Rightist	Mix(Non-Rightist)	Total Voters
President	220	214	434
Vice president	208	226	434
Secretary	246	188	434
Treasurer	236	198	434

From Table **2**, likelihood ratio and Chi-square statistics, as well as AIC and BIC for no class model are very high. Evident of having clusters in the data. Drastic reduction in the values of these measures can be seen for from 1 to 2 class model. In case of AIC, BIC and likelihood ratio statistic the differences are 851.6, 831.2 and 861.6 respectively. Where as, in case of Chi-square statistic the difference is 1492.713 between 2 class and one class model.

Table 2: Results of Fitting Latent Class Models to KUTS Election

	No class	2 class	3 class	4 class
AIC	2402.671	1551.069	1511.16	1517.645
BIC	2418.964	1587.727	1568.183	1595.033
G^2 (Likelihood ratio/deviance statistic)	915.0264	53.42418	3.515523	6.72E-09
X^2 (Chi-square goodness of fit)	1559.74	67.02711	3.572156	6.72E-09
Number of estimated parameters	4	9	14	19
maximum log-likelihood	-1197.34	-766.535	-741.58	-739.823
residual degrees of freedom	11	6	1	-4

N = 434

Best model is one with minimum values of parsimony measures as well as with minimum estimated parameters. Minimum values of AIC and BIC in 3-class is an evidence of an adequate model. Where as, the difference of values of AIC and BIC, from 2-class to 3-class model, are 39.9 and 19.5 respectively. In 3-class model 14 parameters are estimated. This resulted in a single unit for residual degrees of freedom. We will assess conditional item response probabilities for, 2-class (Table **3**) and 3-class (Table **4**), model to affirm the best model.

Table 3: Estimated Class Proportions and Class Conditional Probabilities (with Standard Errors of Estimates) of KUTS Election, for Latent Class Model with Two Classes

	Class 1	Class 2
	52.30% (0.02728107)	**47.70% (0.02728107)**
President		
Rightist	0.9185 (0.024020)	0.0345 (0.018197)
Mix	0.0815 (0.024029)	0.9655 (0.018197)
Vice President		
Rightist	0.8574 (0.027495)	0.0452 (0.017307)
Mix	0.1426 (0.027495)	0.9548 (0.017307)
Secretary		
Rightist	0.9607 (0.016440)	0.1147 (0.028596)
Mix	0.0393 (0.016440)	0.8853 (0.028596)
Treasurer		
Rightist	0.8994 (0.021513)	0.1355 (0.029680)
Mix	0.1006 (0.021513)	0.8645 (0.029680)

In 2-class model (see Table **3**; Figure **1**), maximum vote holders are Rightist group with 52.3% share representing those voters who favor Rightist group. For each post the class conditional probabilities for Rightist group are very high. Where as, remaining 46.56% of the total individuals are in favor of Mix group.

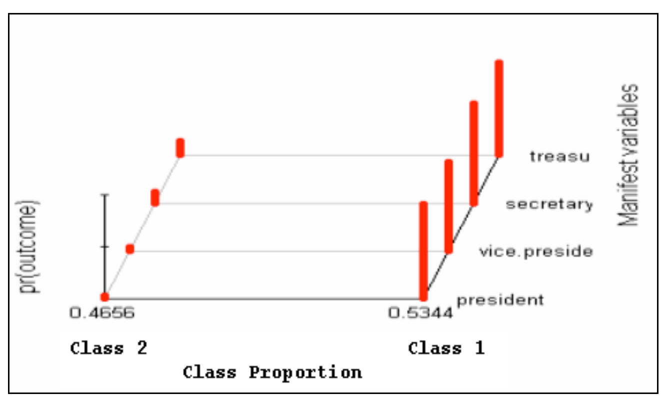

Figure: 1: Estimated class proportions and class conditional probabilities of KUTS election, for the rightist group in 2-class model.

3-class model clearly define three groups of voters. (1) Those who vote members for first two post (president and Vice president) from Mix group and last two (Secretary and Treasurer) from Rightist group (with 22.32% proportion of the total); (2) Voters completely favoring Mix group (with 37.13% proportion of the total); and (3) Voters completely in favor of Rightist group. 40.5% share is of Rightist's voter group. In this class, 100% voters vote for President in favor of Rightist's group. Similarly, voters favor President (with 100% probability) from Mix's group in class 2 (Mix's voter group). Class proportions and class conditional response probabilities are also shown graphically in Figure **2**.

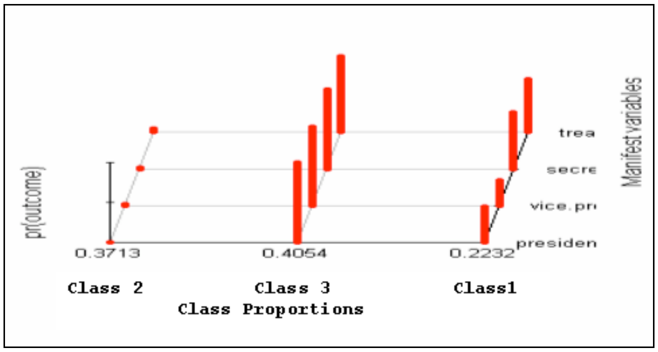

Figure: 2: Estimated class proportions and class conditional probabilities of KUTS election, for the rightist group in 2-class model.

2.2. Parenting Style Data

Diana Baumrind (1966; 1967) developed the theories of parenting styles. She believes that there are three types of parenting styles [19, 20],

1) Authoritarian parenting Style
2) Authoritative parenting Style
3) Permissive parenting Style

Authoritarian parenting is a style, in which parents are more demanding but non responsive. This style is also called strict parenting. These types of parents have high expectations, regarding their defined rules and restrictions, of compliance. Children experiencing such style are often complaining generation-gap problem. As, parents do not (allow open discussions) listen to them instead impose decisions without explaining the reasons. Children rear from such parenting style may results in low social inclination, break down and runaway.

In Authoritative parenting style, parents are demanding as well as responsive, also called balanced parenting. These types of parents give room and encourage their children to share his/her point of view with having control on their actions. They are concern about the needs of their children. They set clear standards allowing children self-will and disciplined conformity. Children experiencing authoritative parenting style are supposed to have a higher self esteem and independence.

Responsive parents places few demands of responsibility on children lead to permissive parenting style. They have little control on them besides they present them as a resource of fulfilling their needs and wishes. Such parenting may result in spoiled children. Children may tend to engage in misconduct as they want everything to be done in a way they like and have no control on their behavior. In some cases, children mature quickly and are able to live independent life without some ones help.

We want to study the effect of different parenting style on the self esteem of children. The variables are taken from a survey conducted through Parental Authority questionnaire (PAQ) designed to measure parental authority, or disciplinary practices, from point of view of the child of any age [18]. Along with parenting style and self esteem we also include gender and thoughts of suicide in the analysis. Coding of these variables is given in Table **5**. From Table **6**, AIC suggest that the model fitted is a 3-class model. Where, BIC is at its minimum in 2-class model. We will be discussing class item conditional probabilities for both 2-class and 3-class model.

Table 5: Codes Assign to Each Manifest Variable to Parenting Style Data

	Variables			
Code Assign	**Gender**	**Parenting Style**	**Suicidal Ideation**	**Self Esteem**
1	Male	Authoritarian	Low	Low
2	Female	Authoritative	Mild-Moderate	High
3	---	Permissive	Severe	---

Table 6: Results of Fitting Latent Class Models to Parenting Style Data

	No Latent Class	2 Latent Class	3 Latent Class	4 Latent Class
AIC	3652.099	3402.857	3390.89	3402.028
BIC	3678.725	3460.548	3479.645	3521.847
G^2	312.9928	49.75075	23.78365	20.92207
X^2	418.2356	46.92199	20.98417	18.60521
Maximum log-likelihood	-1820.05	-1688.43	-1675.45	-1674.01
number of estimated parameters	6	13	15	27

N=625

Class item response conditional probabilities for 2-class model are presented in Table **7**; standard errors of estimates are also given in parenthesis. Class 1 in 2-class model represent males with almost 52% probability, who believe that they are experiencing Authoritative and Permissive parenting style. So, they are some what relaxed in taking decisions and feel free. With a very low suicidal Ideation a quite large number of individuals in this group believe that they have high Self Esteem. 83.24% people belong to this class.

Table 7: Estimated Class Proportions and Class conditional probabilities (with Standard Errors of Estimates) of Parenting Style Data, for Latent Class Model with Two Classes

Indicators	Class 1	Class 2
	83.24% (0.0276567)	**16.76% (0.0276567)**
Gender		
Male	0.5192 (0.02243743)	0.3809 (0.05098442)
Female	0.4808 (0.02243743)	0.6191 (0.05098442)
Parenting Style		
Authoritarian	0.2794 (0.02123919)	0.6265 (0.05204465)
Authoritative	0.3689 (0.02171427)	0.1724 (0.04118376)
Permissive	0.3516 (0.02129322)	0.2011 (0.04057631)
Suicidal Ideation		
Low	0.9892 (0.01244069)	0.0037 (0.12330489)
Mild-Moderate	0.0108 (0.01244069)	0.7863 (0.10548435)
Severe	0.000 (0.000)	0.21 (0.04922041)
Self Esteem		
Low	0.1295 (0.02411008)	0.8747 (0.05461173)
High	0.8705 (0.02411008)	0.1253 (0.05461173)

Class 2 in 2-class model represent more to females than males who believes that they have a very low Self Esteem and they have thoughts of Suicide of Mild-Moderate intensity. This may be due to Authoritarian parenting style they experienced (with almost 63% probability).

In 3-class model (see Table **8** and Figure **3**), conditional item response probabilities show that class with 15.4% share, is a mix class with more females who experienced authoritarian parenting style and have very low self esteem with a high probability of mild-moderate chance of suicidal ideation.

The other two classes, in 3-class model, are very much clear. Classes with 43.1% and 41.5% class proportions, represent males and females respectively. Individual in both the groups have a very high self esteem with a very low chances of having thoughts of suicide. As well as, they experienced authoritative and permissive parenting style more as compared to authoritarian parenting style.

Close analysis of class item probabilities reveals that both, the 2-class model and 3-class model, give same results. In 3-class model class 2 and 3 (see Table: **8**) are the sub groups of class 1 (see Table: **7**) of 2-class model.

CONCLUSION

In KUTS election data manifest variables are binary. Whereas, parenting style data include manifest variable having two or three responses. Interpretation of models for both data sets is therefore a little bit different. It goes from a relatively simple to a bit difficult analysis of fitted model. Moreover, respondents of KUTS election data were those mature people who were in their professional life and were aware and able to choose

Table 8: Estimated Class Proportions and Class Conditional Probabilities (with Standard Errors of Estimates) of Parenting Style Data, for Latent Class Model with Three Classes

Indicators	Class 1	Class 2	Class 3
	15.36% (0.020820)	**43.10% (0.016517)**	**41.50% (0.0199540)**
Sex			
Male	0.42299 (0.0547)	1.000(0.0000)	0.000 (0.000)
Female	0.5770 (0.0547)	0.000 (0.0000)	1.000 (0.000)
Parenting Style			
Authoritarian	0.6386 (0.05242832)	0.28985 (0.02774)	0.27582 (0.02848)
Authoritative	0.1593 (0.04048331)	0.37231 (0.02953)	0.36366 (0.03026)
Permissive	0.2020 (0.04275375)	0.33783 (0.02883)	0.36051 (0.03015)
Suicidal Ideation			
Low	0.000 (0.0000)	0.99855 (0.006040)	0.94760 (0.02061)
Mild-Moderate	0.77085 (0.04593)	0.00144 (0.006039)	0.05239 (0.02061)
Severe	0.22914 (0.04593)	0.000 (0.000)	0.000 (0.000)
Self Esteem			
Low	0.93319 (0.0457)	0.07063 (0.01567)	0.19405 (0.02530)
High	0.06680 (0.04575)	0.92936 (0.01567)	0.80594 (0.02530)

between what is good or bad things for them. The KUTS election data is about the respondent's decision that they took in judging others. Where as, respondents of parenting style data are those teenagers who are preparing for professional life. Responses were their perception the way they are and the way they have being brought up. The purpose of using completely two different data sets is to acknowledge the aspect of Latent structure model. Application of this technique enables us to identify the prevailing groups, for both the data set, whether the responses are perceptions or judgment. In case of KUTS election data 3-class model gives better results. Where as, in parenting style data, BIC is at its minimum at 2-class model but conditional item response probabilities of 3-class model give clear grouping in accordance to gender.

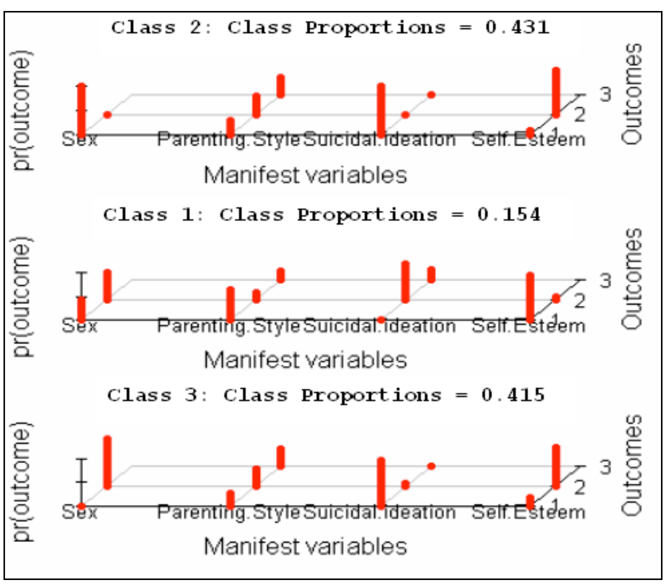

Figure: 3: Estimated class proportions and class conditional probabilities of 3-class model for parenting style data.

ACKNOWLEDGEMENT

We would like to thank Dr. Zahid Iqbal (Assistant Professor, International Islamic University) for providing parenting style data.

REFERENCES

[1] Anderson TW. An introduction to multivariate statistical analysis. 2nd ed. John Wiley: New York; 1984.

[2] Lazarsfeld PF, Henry NW. Latent structure analysis. Boston: Houghton Mi_in; 1968.

[3] Bartholomew DJ. Factor analysis for categorical data. Journal of the Royal Statistical Society, Series B (Methodological) 1980; 42(3): 293-321.

[4] Goodman LA, Kruskal WH. Measures of association for cross classifications. J. Amer. Statist. Assoc. 1954; 49(268): 732-764.
http://dx.doi.org/10.2307/2281536

[5] Clogg CC, Goodman LA. Latent structural analysis of a set of multidimensional contingency tables. J. Amer. Statist. Assoc 1984; 79: 762-71.
http://dx.doi.org/10.2307/2288706

[6] Clogg CC. Some latent structure models for analysis of likert-type data. Social Science Research 1979; 8: 287-301.
http://dx.doi.org/10.1016/S0049-089X(79)80001-0

[7] McCutcheon AL. Latent class analysis. Newbury Park: Sage Publications; 1987.

[8] Vermunt JK. Multilevel latent class models. Sociological Methodology 2003; 33: 213-39. http://dx.doi.org/10.1111/j.0081-1750.2003.t01-1-00131.x

[9] Muthen BO. Beyond SEM: General latent variable modeling. Behaviormetrika 2002; 29(1): 81-117. http://dx.doi.org/10.2333/bhmk.29.81

[10] Agresti A. Categorical data analysis. 2nd ed. New York: Wiley; 2002. http://dx.doi.org/10.1002/0471249688

[11] Dempster AP, Laird NM, et al. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society, Series B (Methodological) 1977; 39 (1): 1-38.

[12] Linzer DA, Lewis J. poLCA: An R package for polytomous variable latent class analysis. Journal of Statistical Software 2011; 42(10): 1-29.

[13] Linzer DA, Lewis J. poLCA: Polytomous variable latent class analysis. R package version 1.3, 2011. (URL http://CRAN.R-project.org/package=poLCA).

[14] R Development Core Team 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[15] Goodman L. The multivariate analysis of qualitative data: Interactions among Multiple Classifications. J. Amer. Statist. Assoc 1970; 65: 226-56. http://dx.doi.org/10.2307/2283589

[16] Akaike H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267-281 in B. N. Petrov and F. Csaki, editors. Second international symposium on information theory. Akademiai Kiado, Budapest, Hungary.

[17] Schwarz GE. Estimating the dimension of a model. Annals of Statistics 1978; 6(2): 461–4. http://dx.doi.org/10.1214/aos/1176344136

[18] Buri, J.R. 1991. Parental authority questionnaire. University of St. Thomas, 2115 Simmit Avenue, St. Paul, MN 55105.

[19] Baumrind D. Effects of authoritative parental control on child behavior, Child Development 1966; 37(4): 887-907. http://dx.doi.org/10.2307/1126611

[20] Baumrind, D. Child care practices anteceding three patterns of preschool behavior. Genetic Psychology Monographs. 1967; 75(1): 43-88.

Received on 24-10-2011 Accepted on 08-12-2011 Published on 13-02-2012

DOI: http://dx.doi.org/10.6000/1927-5129.2012.08.01.14