A Nontechnical Introduction to Latent Class Models

by

Jay Magidson, Ph.D.
Statistical Innovations Inc.

Jeroen K. Vermunt, Ph.D.
Tilburg University, the Netherlands

Over the past several years more significant books have been published on latent class (LC) and finite mixture models than any other class of statistical models. The recent increase in interest in LC models is due to the development of extended computer algorithms, which allow today's computers to perform latent class analysis on data containing more than just a few variables. In addition, researchers are realizing that the use of latent class models can yield powerful improvements over traditional approaches to cluster, factor, regression/segmentation and neural network applications, and related graphical displays.

What are Latent Class Models?

Traditional models used in regression, discriminant and log-linear analysis contain parameters that describe only relationships between the observed variables. LC models (also known as finite mixture models) differ from these by including one or more discrete unobserved variables. In the context of marketing research, one will typically interpret the categories of these latent variables, the latent classes, as clusters or segments (Dillon and Kumar 1994; Wedel and Kamakura 1998). Among other uses, LC analysis provides a powerful new tool to identify important market segments in target marketing.

Recently, the close connection between LC models and random effects models (REM) has been made apparent (Vermunt and Van Dijk, 2002; Agresti, 2002). In addition, the close connection between latent classes and "hidden layer nodes" in the most widely used neural net model, the multilayer perceptron (MLP) has been clarified (Vermunt and Magidson, 2002).. These recent developmentsopen the door to the use of latent class models for nonlinear regression applications, providing improvements over the current approaches to both REM and MLP in speed and efficiency of estimation, as well as in interpretation of the results.

LC models do not rely on the traditional modeling assumptions which are often violated in practice (linear relationship, normal distribution, homogeneity). Hence, they are less subject to biases associated with data not conforming to model assumptions. Also, for improved cluster or segment description (and prediction), the relationship between the latent classes and external variables (covariates) can be assessed simultaneously with the identification of the classes (clusters, segments). This eliminates the need for the usual

second stage of analysis where a discriminant analysis is performed to relate the resulting clusters or factors obtained from a traditional cluster or factor analysis to demographic and other variables. In addition, LC models have recently been extended (Vermunt and Magidson, 2000, 2002) to include variables of mixed scale types (nominal, ordinal, continuous and/or count variables) in the same analysis.

Kinds of Latent Class Models

Three common statistical application areas of LC analysis are those that involve
1) clustering of cases,
2) variable reduction and scale construction, and
3) prediction of a dependent variable.

This paper introduces the three major kinds of LC models:
- LC Cluster Models,
- LC Factor Models,
- LC Regression and Choice Models.

This paper also describes the use of LC models as an alternative to the neural network approach of modeling nonlinearities. Our illustrative examples make use of the Latent GOLD® computer program (Vermunt and Magidson, 2000).

LC Cluster Models

The **LC Cluster** model:
- identifies *clusters* which group together *persons* (cases) who share similar interests/values/characteristics/behavior,
- includes a K-category latent variable, each category representing a cluster.

Advantages over traditional types of cluster analysis include:
- probability-based classification: Cases are classified into clusters based upon membership probabilities estimated directly from the model,
- variables may be continuous, categorical (nominal or ordinal), or counts or any combination of these,
- demographics and other covariates can be used for cluster description.

Typical marketing applications include:
- exploratory data analysis,
- development of behavioral based and other segmentations of customers and prospects.

Traditional clustering approaches utilize "unsupervised" learning/classification algorithms that group cases that are "near" each other according to some ad hoc definition of "distance". In the last decade interest has shifted towards model-based approaches which use estimated membership probabilities to classify cases into the appropriate cluster. The most popular model-based approach is known as mixture-model clustering, where each latent class represents a hidden cluster (McLachlan and Basford, 1988, Vermunt and Magidson, 2002a). Within the marketing research field, this method is sometimes referred to as "latent discriminant analysis" (Dillon and Mulani, 1999). Today's high-speed computers make these computationally intensive methods practical.

In the case of continuous variables, Magidson and Vermunt (2002) show that LC clustering outperforms the traditional K-means algorithm. For the general finite mixture model, not only continuous variables, but also variables that are ordinal, nominal or counts, or any combination of these can be included. Also, covariates can be included for improved cluster description.

As an example, we used the LC cluster model to develop a segmentation of current bank customers based upon the types of accounts they have. Separate models were developed, each specifying a different number of clusters. The model selected was the one that had the lowest BIC statistic.

This criteria resulted in 4 segments which were named:
1) Value Seekers (15% of customers),
2) Conservative Savers (35% of customers),
3) Mainstreamers (40% of customers),
4) Investors (10% of customers).

For each customer, the model gave estimated membership probabilities for each segment based on their account mix. The resulting segments were verified to be very homogeneous and to differ substantially from each other not only with respect to their mix of accounts, but also with respect to demographics, and profitability. In addition, examination of survey data among the sample of customers for which customer satisfaction data were obtained found some important attitudinal and satisfaction differences between the segments as well. Value seekers were youngest and a high percentage were new customers. Basic savers were oldest.

Investors were the most profitable customer segment by far. Although only 10% of all customers, they accounted for over 30% of the bank's deposits. Survey data pinpointed the areas of the bank with which this segment was least satisfied and a LC regression model (see below) on follow-up data was used to relate their dissatisfaction to attrition. The primary uses of the survey data was to identify reasons for low satisfaction and to develop strategies of improving satisfaction in the manner that increased retention.

This methodology of segmenting based on behavioral information available on all customers offers many advantages over the common practice of developing segments from survey data and then attempting to allocate all customers to the different clusters.

Advantages of developing a segmentation based on *behavioral* data include:
- past behavior is known to be the best predictor of future behavior,
- all customers can be assigned to a segment directly, not just the sample for which survey data is available,
- improved reliability over segmentations based on attitudes, demographics, purchase intent and other survey variables. When segment membership is based on survey data, a large amount of classification error is almost always present for non-surveyed customers) .

LC Factor Models

The **LC Factor** model:
- identifies *factors* which group together *variables* sharing a common source of variation,
- can include several ordinal latent factors, each of which contains 2 or more levels,
- is similar to maximum likelihood (ML) factor analysis in that its use may be exploratory or confirmatory and factors may be assumed to be correlated or uncorrelated (orthogonal).

Advantages over traditional factor analysis are:
- factors need not be rotated to be interpretable,
- ML estimates for factor scores are obtained directly from the model without imposing additional assumptions,
- variables may be continuous, categorical (nominal or ordinal), or counts or any combination of these,
- extended factor models can be estimated that include covariates and correlated residuals.

 Typical marketing applications include:
- development of composite variables from attitudinal survey items,
- development of perceptual maps and other kinds of biplots which relate product and brand usage to behavioral and attitudinal measures and to demographics,
- estimation of factor scores,
- direct conversion from factors to segments.

The conversion of ordinal factors to segments is straightforward. For example, consider a model containing 2 dichotomous factors. In this case, the LC factor model provides ML estimates for the membership classification probabilities directly for 4 clusters (segments) based on the classification of cases as high vs. low on each factor: segment 1 = (low, low); segment 2 = (low, high); segment 3 = (high, low) and segment 4 = (high, high). Magidson and Vermunt (2001) found that LC factor models specifying

uncorrelated factors often fit data better than comparable LC cluster models (i.e., cluster models containing the same number of parameters).

Figure 1 provides a bi-plot in 2-factor space of lifestyle interests where the horizontal axis represents the probability of being high on factor 1 and the vertical axis the probability of being high on factor 2. The variable AGE was included directly in the LC factor model as a covariate and therefore shows up in the bi-plot to assist in understanding the meaning of the factors. For example, we see that persons aged 65+ are most likely to be in the (low, high) segment, as are persons expressing an interest in sewing. As a group, their (mean) factor scores are (Factor 1, Factor 2) = (.06, .67).

[INSERT FIGURE 1 ABOUT HERE]

Since these factor scores have a distinct probabilistic interpretation, this bi-plot represents an improvement over traditional biplots and perceptual maps (see Magidson and Vermunt 2001). Individual cases can also be plotted based on their factor scores.

In addition, LC factor analysis can be performed using fewer variables than traditional factor analysis. In traditional factor analysis, at least 3 variables are required which must be continuous, and such an analysis of 3 variables can identify only a single factor. With LC factor analysis, 3 dichotomous variables similarly yield 1 factor. However, LC models are not limited to dichotomies, and the inclusion of covariates can allow additional factors to be identified. For example, the analysis of 1 or 2 continuous variables, even without covariates, may yield a 2 (or more)-factor solution. Moreover, results from a 2-factor solution can be presented in a bi-plot display.

The factor model can also be used to deal with measurement and classification errors in categorical variables. It is actually equivalent to a latent trait (IRT) model without the requirement that the traits be normally distributed. For further details on the LC factor model, see Magidson and Vermunt (2001, 2003).

LC Regression Models

The **LC Regression** model, also known as the **LC Segmentation** model:
- is used to predict a dependent variable as a function of predictors,
- includes an R-category latent variable, each category representing a homogeneous population (class, segment),
- different regressions are estimated for each population (for each latent segment),
- classifies cases into segments and develops regression models simultaneously.

Advantages over traditional regression models include:
- relaxing the traditional assumption that the same model holds for all cases (R=1) allows the development of separate regressions to be used to target each segment,
- diagnostic statistics are available to determine the value for R,

- for R > 1, covariates can be included in the model to improve classification of each case into the most likely segment.

Typical marketing applications include:
- customer satisfaction studies: identify particular determinants of customer satisfaction that are appropriate for each customer segment,
- conjoint studies: identify the mix of product attributes that appeal to different market segments,
- more generally: identify latent segments that can explain unobserved heterogeneity in the data.

With respect to conjoint studies, LC models have been utilized with rating models, and more recently with discrete choice and ranking models. Discrete choice modeling allows estimation of consumer utilities, a theoretical development that led to the award of a Nobel prize (McFadden, 1986). LC choice models provide an important advance in these models, allowing for different utilities to be estimated for each latent segment (McFadden and Train, 2000, Greene and Hensher, 2002, Vermunt and Magidson, 2003).

Like traditional regression modeling, LC regression requires a computer program. As LC regression modeling is relatively new, very few programs currently exist. Our comparisons between LC regression and traditional linear regression are based on the particular forms of LC regression that are implemented in the Latent GOLD® and Latent GOLD Choice (Vermunt and Magidson, 2003) programs. For other software see Wedel and DeSarbo (1994),and Wedel and Kamakura (1998).

Typical regression programs utilize ordinary least squares estimation in conjunction with a linear model. In particular, such programs are based on two restrictive assumptions about data that are often violated in practice:
1) the dependent variable is continuous with prediction error normally distributed,
2) the population is homogeneous - one model holds for all cases.

LC regression as implemented in the Latent GOLD® program relaxes these assumptions:
1) it accommodates dependent variables that are continuous, categorical (binary, polytomous nominal or ordinal), binomial counts, or Poisson counts,
2) the population needs not be homogeneous (i.e., there may be multiple populations as determined by the BIC statistic).

One potential drawback for LC models is that there is no guarantee that the solution will be the maximum likelihood solution. LC computer programs typically employ the EM or Newton Raphson algorithm which may converge to a local as opposed to a global maximum. Some programs provide randomized starting values to allow users to increase the likelihood of converging to a global solution by starting the algorithm at different randomly generated starting places.

An additional approach is to use Bayesian prior information in conjunction with randomized starting values which eliminates the possibility of obtaining boundary

(extreme) solutions and reduces the chance of obtaining local solutions. Generally speaking, we have achieved good results using 10 randomized starting values and small Bayes constants (the default option in the Latent GOLD program). For further discussion of these issues and a comparison of Latent GOLD with some other programs see Uebersax (2000).

In addition to using predictors to estimate separate regression model for each class, covariates can be specified to refine class descriptions and improve classification of cases into the appropriate latent classes. Typically, LC regression analysis consists of 4 simultaneous steps:
1) identify latent classes or hidden segments
2) use demographic and other covariates to predict class membership
3) classify cases into the appropriate classes/segments, and
4) estimate regression models for each class

While LC regression analysis is generally conducted with dependent variables consisting of a single measurement per observation, dependent variables may also include repeated measures (correlated observations) over time, or repeated ratings of the kind often collected in conjoint marketing studies where each person is asked to rate different brands or different attributes of a brand. Below is an example of a full factorial conjoint study designed to assist in the determination of the mix of product attributes for a new product.

Conjoint Case Study

In this example, 400 persons were asked to rate each of 8 different attribute combinations regarding their likelihood to purchase. Hence, there are 8 records per case; one record for each cell in this 2x2x2 conjoint design based on the following attributes:
- *FASHION (1 = Traditional; 2 = Modern),*
- *QUALITY (1 = Low; 2 = High),*
- *PRICE (1 = Lower; 2 = Higher) .*

The dependent variable (*RATING*) is the rating of purchase intent on a five-point scale. The three attributes listed above are used as predictor variables in the model and the following demographic variables are used as covariates:
- *SEX (1 = Male; 2 = Female),*
- *AGE (1 = 16-24; 2 = 25-39; 3 = 40+).*

The goal of a traditional conjoint study of this kind is to determine the relative effects of each attribute in influencing one's purchase decision; a goal attained by estimating regression (or logit) coefficients for these attributes. When the LC regression model is used with the same data, a more general goal is attained. First, it is determined whether the population is homogeneous or whether there exists two or more distinct populations (latent segments) which differ with respect to the relative importance placed on each of the three attributes. If multiple segments are found, separate regression models are

estimated simultaneously for each. For example, for one segment, price may be found to influence the purchase decision, while a second segment may be price insensitive, but influenced by quality and modern appearance.

We will treat *RATING* as an *ordinal* dependent variable and estimate several different models to determine the number of segments (latent classes). We will then show how this methodology can be used to describe the demographic differences between these segments and to classify each respondent into the segment which is most appropriate.

We estimated one- to four-class models with and without covariates. Table 1 reports the obtained test results. The BIC values indicate that the three-class model is the best model (BIC is lowest for this model) and that the inclusion of covariates significantly improves the model.

[INSERT TABLE 1 ABOUT HERE]

The parameter estimates of the three-class model with covariates are reported in Tables 2 and 3 and 4. As can be seen from the first row of Table 2, segment 1 contains about 50% of the subjects, segment 2 contains about 25% and segment 3 contains the remaining 25%. Examination of class-specific probabilities shows that overall, segment 1 is least likely to buy (only 5% are *Very Likely* to buy) and segment 3 is most likely (21% are *Very Likely* to buy).

[INSERT TABLES 2 and 3 ABOUT HERE]

The beta parameter for each predictor is a measure of the influence of that predictor on *RATING*. The beta effect estimates under the column labeled Class 1 suggest that segment 1 is influenced in a positive way by products for which *FASHION = Modern* (beta = 1.97) and in negative way by *PRICE = Higher* (beta = -1.04), but not by *QUALITY* (beta is approximately 0). We also see that segment 2 is influenced by all 3 attributes, having a preference for those product choices that are *modern* (beta = 1.14), *high quality* (beta = .85) and *lower priced* (beta = -0.99). Members of segment 3 prefer *high quality* (beta = 2.06) and the *lower* (beta = -.94) product choices, but are not influenced by *FASHION*.

Note that *PRICE* has more or less the same influence on all three segments. The Wald (=) statistic indicates that the differences in these beta effects across classes are not significant (the *p*-value = .68 which is much higher than .05, the standard level for assessing statistical significance). This means that all 3 segments exhibit price sensitivity to the same degree. This is confirmed when we estimate a model in which this effect is specified to be class-independent. The *p*-value for the Wald statistic for *PRICE* is 2.9×10^{-107} indicating that the amount of price sensitivity is highly significant.

With respect to the effect of the other two attributes we find large between-segment differences. The predictor *FASHION* has a strong influence on segment 1, a less strong effect on segment 2, and virtually no effect on segment 3. *QUALITY* has a strong effect

on segment 3, a less strong effect on segment 2, and virtually no effect on segment 1. The fact that the influence of *FASHION* and *QUALITY* differs significantly between the 3 segments is confirmed by the significant *p*-values associated with the Wald(=) statistics for these attributes. For example, for *FASHION*, the *p*-value = 3.0×10^{-42}.

The beta parameters of the regression model can be used to name the latent segments. Segment 1 could be named the "Fashion-Oriented" segment, segment 3 the "Quality-Oriented" segment, and segment 2 is the segment that takes into account all 3 attributes in their purchase decision.

[INSERT TABLE 4 ABOUT HERE]

The parameters of the (multinomial logit) model for the latent distribution appear in Table 4. These show that females have a higher probability of belonging to the "Fashion-oriented" segment (segment 1), while males more often belong to segment 2. The Age effects show that the youngest age group is over-represented in the "Fashion-oriented" segment, while the oldest age group is over-represented in the "Quality oriented" Segment.

Applying Structure to Neural Network Models

While current applications of neural networksmaintain a "black-box-like" nature, recent advances by statisticians promise similar LC applicationsin the near future that provide more efficient and speedier estimation, and more easily interpretable results (see Kay and Titterington, 1999, Vermunt and Magidson, 2002b).

For example, consider the case of nonlinear response models where say response or net revenue from a direct marketing effort is the dependent variable, a "supervised" learning situation. In this setting, neural network models have increasingly been employed in an attempt to incorporate nonlinear relationships in the model. Such nonlinearities are typically included in a neural net model by including 2 or more nodes in a "hidden layer" of a multi-layer perceptron (MLP). However, these models can be structured as LC models by taking each node to be a latent factor (Vermunt and Magidson, 2002b). By doing such, not only do the parameter estimates become more meaningful, but graphical displays can visually display the nonlinearities.

(INSERT FIGURE 2 ABOUT HERE)

For example, Figure 2 shows results of using a 2-factor MLP-like LC model to predict the frequency of Internet Usage as a nonlinear function of age, education and gender. Frequency was measured by the number of times used during the previous week (Source: 1999 Mediamark Research Inc. Survey of American Consumers). From this display, we see the following nonlinear relationships:

- Two distinct dimensions of Internet Usage are evident:

- Factor 1 distinguishers non-users (0 frequency) from users (>0 frequency)
- Factor 2 distinguishes heavy from light users.

- AGE is primarily predictive of Factor 1, EDUCATION of Factor 2
  - AGE -- Older users, especially those aged 65+ are likely to be non-users.
  - EDUCATION – The more educated, especially those with at least a masters degree, use the internet most frequently.

The points displayed in Figure 2 are plotted at the posterior probabilities (p1,p2) where p1 represents the conditional probability of being high on Factor 1 and p2 represents the conditional probability of being high on Factor 2. By connecting the points associated with the categories of the same variable, Figure 2 clearly depicts the primary nonlinearities in the data.

Conclusions

We introduced three kinds of LC models and described applications of each that are of interest in marketing research, survey analysis and related fields. It was seen that LC analysis can be used as a replacement for traditional cluster analysis techniques, as a factor analytic tool for reducing dimensionality, as a tool for estimating separate regression models for each segment, and as an alternative to neural net approach for modeling nonlinearities in data. Overall, these models offer powerful new approaches for identifying market segments and for many other applications of interest to marketing researchers.

BIOS

Jay Magidson is founder and president of Statistical Innovations, a Boston based consulting, training and software development firm specializing in segmentation modeling. His clients have included A.C. Nielsen, Household Finance, and National Geographic Society. He is widely published on the theory and applications of multivariate statistical methods, and was awarded a patent for a new innovative graphical approach for analysis of categorical data. He taught statistics at Tufts and Boston University, and is chair of the Statistical Modeling Week workshop series. Dr. Magidson designed the SPSS CHAID™ and GOLDMineR® programs, and is the co-developer (with Jeroen Vermunt) of the Latent GOLD® Latent GOLD Choice™ programs.

Jeroen Vermunt is Professor in the Department of Methodology and Statistics of the Faculty of Social and Behavioral Sciences, and Research Associate at the Work and Organization Research Center at Tilburg University in the Netherlands. He has taught a variety of courses and seminars on log-linear analysis, latent class analysis, item response models, models for non-response, and event history analysis all over the world, as well as published extensively on these subjects. Professor Vermunt is developer of the LEM program and co-developer (with Jay Magidson) of the Latent GOLD® and Latent GOLD Choice™ programs.

References

Agresti, A. (2002). *Categorical Data Analysis*. Second Edition. New York: Wiley.

Dillon, W.R., and Kumar, A. (1994). Latent structure and other mixture models in marketing: An integrative survey and overview, chapter 9 in R.P. Bagozzi (ed.), *Advanced methods of Marketing Research*, 352-388,Cambridge: Blackwell Publishers.

Dillon, W.R.. and Mulani, N. (1989) LADI: A latent discriminant model for analyzing marketing research data. *Journal of Marketing Research*, 26, 15-29.

Greene, W.H. and D.A. Hensher, (2002) "A latent class model for discrete choice analysis: Contrasts with mixed logit", working paper ITS-WP-02-08, Institute of Transport studies, The Australian Key Centre in Transport Management, The University of Sydney and Monash University.

Kay and Titterington (Eds.) 1999, *Statistics and Neural Networks: Advances at the Interface*. Oxford: Oxford University Press.

Magidson J., and Vermunt, J.K. (2001), *Latent Class Factor and Cluster Models, Bi-plots and Related Graphical Displays*.Chapter 5 in Becker and Sobel (Eds.) Sociological Methodology, Vol. 31, 223-264. .

Magidson J. and Vermunt, J.K. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20, 37-44.

Magidson J., and Vermunt, J.K. (2003) "Comparing Latent Class Factor Analysis with the Traditional Approach in Datamining", forthcoming in Statistical Applications of Datamining.

McFadden, D. (1986) "The choice theory approach to marketing research", Marketing Science 5(4):275-97.

McFadden, D. and Train (2000), "Mixed MNL models for discrete response", Journal of Applied Econometrics, 15, 447-470.

McLachlan, G.J., and Basford, K.E. (1988). *Mixture models: inference and application to clustering*. New York: Marcel Dekker.

Uebersax (2000) "A brief study of local maximum solutions in latent class analysis", http://ourworld.compuserve.com/homepages/jsuebersax/local.htm.

Vermunt, J.K. & Magidson, J. (2000). *Latent GOLD 2.0 User's Guide*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K. & Magidson, J. (2002a). "Latent Class Cluster Analysis", chapter 3 in J.A. Hagenaars and A.L. McCutcheon (eds.), *Advances in Latent Class Analysis*. Cambridge University Press.

Vermunt, J.K. and Magidson J., (2002b), "Latent Class Models for Classification", Computational Statistics and Data Analysis, Elsevior (in press).

Vermunt, J.K. & Magidson, J. (2003, forthcoming). *Latent GOLD Choice 3.0 User's Guide*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K. and Van Dijk. L. (2001). A nonparametric random-coefficients approach: the latent class regression model. *Multilevel Modelling Newsletter*, 13, 6-13.

Wedel, M., and DeSarbo, W.S (1994). A review of recent developments in latent class regression models. R.P. Bagozzi (ed.), *Advanced methods of Marketing Research*, 352-388, Cambridge: Blackwell Publishers.

Wedel, M., and Kamakura, W.A. (1998). *Market segmentation: Concepts and methodological foundations*. Boston: Kluwer Academic Publishers.

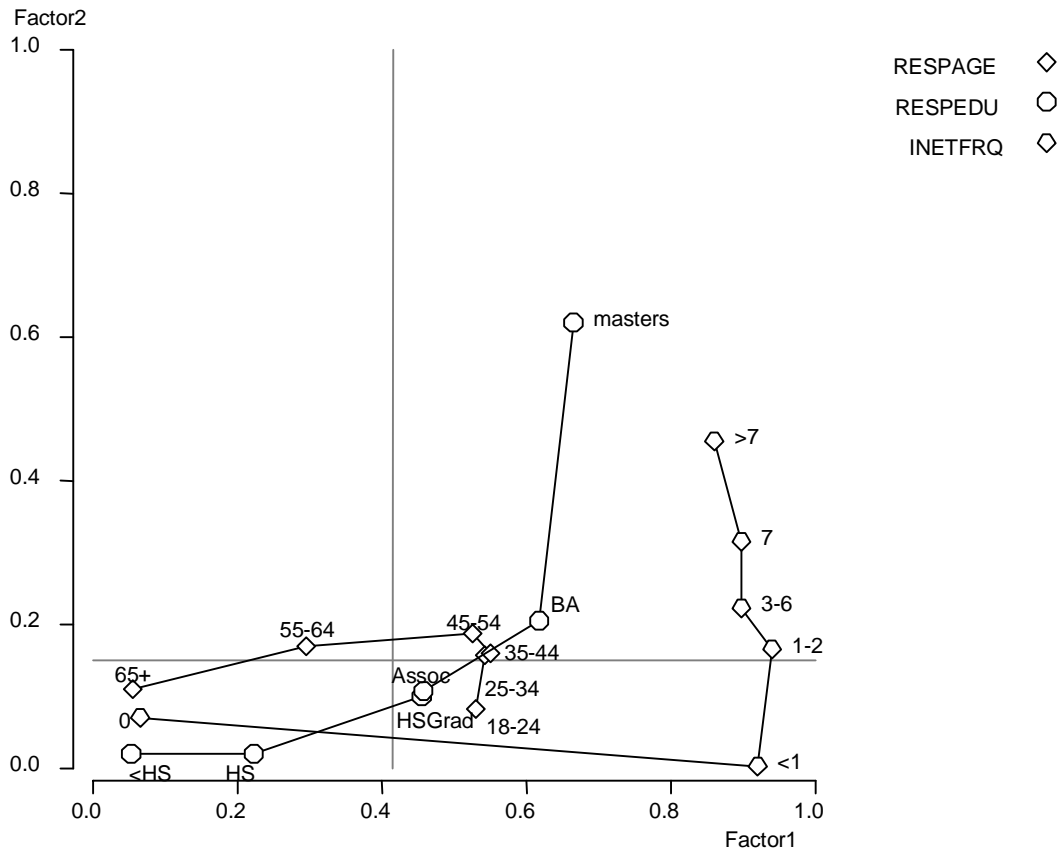**Figure 1: Bi-plot for life-style data (Source: The Polk Co.)**
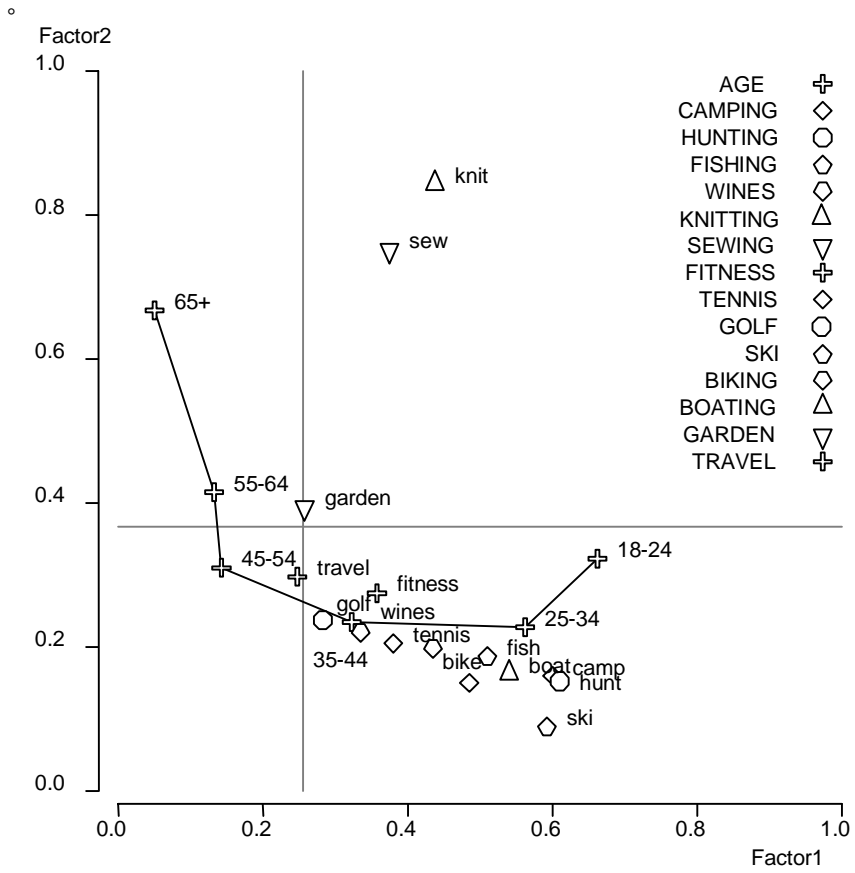
Figure 2: Bi-plot for Internet Frequency-of-Use data (Source: Mediamark Research Inc.)

Table 1: Test results for regression models for conjoint data

Model	Log-likelihood	BIC-value	Number of parameters
Without covariates			
One segment	-4402	8846	7
Two segments	-4141	8319	15
Three segments	-4087	8312	23
Four segments	-4080	8346	31
With covariates			
Two segments	-4088	8284	18
Three segments	-4036	8246	29
Four segments	-4026	8293	40

♦ *Table 2: Profile output*

	Class 1	Class 2	Class 3
Segment Size	0.49	0.26	0.25
Rating			
Very Unlikely	0.21	0.10	0.05
Not Very Likely	0.43	0.20	0.12
Neutral	0.20	0.37	0.20
Somewhat Likely	0.10	0.21	0.43
Very Likely	0.05	0.11	0.21

♦ *Table 3: Beta's or parameters of model for dependent variable*

	Class 1	Class 2	Class 3	Wald	p-value	Wald(=)	p-value
Fashion	1.97	1.14	0.04	440.19	4.4e-95	191.21	3.0e-42
Quality	0.04	0.85	2.06	176.00	6.5e-38	132.33	1.8e-29
Price	-1.04	-0.99	-0.94	496.38	2.9e-107	0.76	0.68

♦ *Table 4: Gamma's: parameters of model for latent distribution*

	Class 1	Class 2	Class 3	Wald	p-value
Sex					
Male	-0.56	0.71	-0.15	24.47	4.9e-6
Female	0.56	-0.71	0.15		
Age					
16-25	0.84	-0.59	-0.24	53.09	8.1e-11
26-40	-0.32	0.59	-0.27		
40+	-0.52	0.01	0.51		