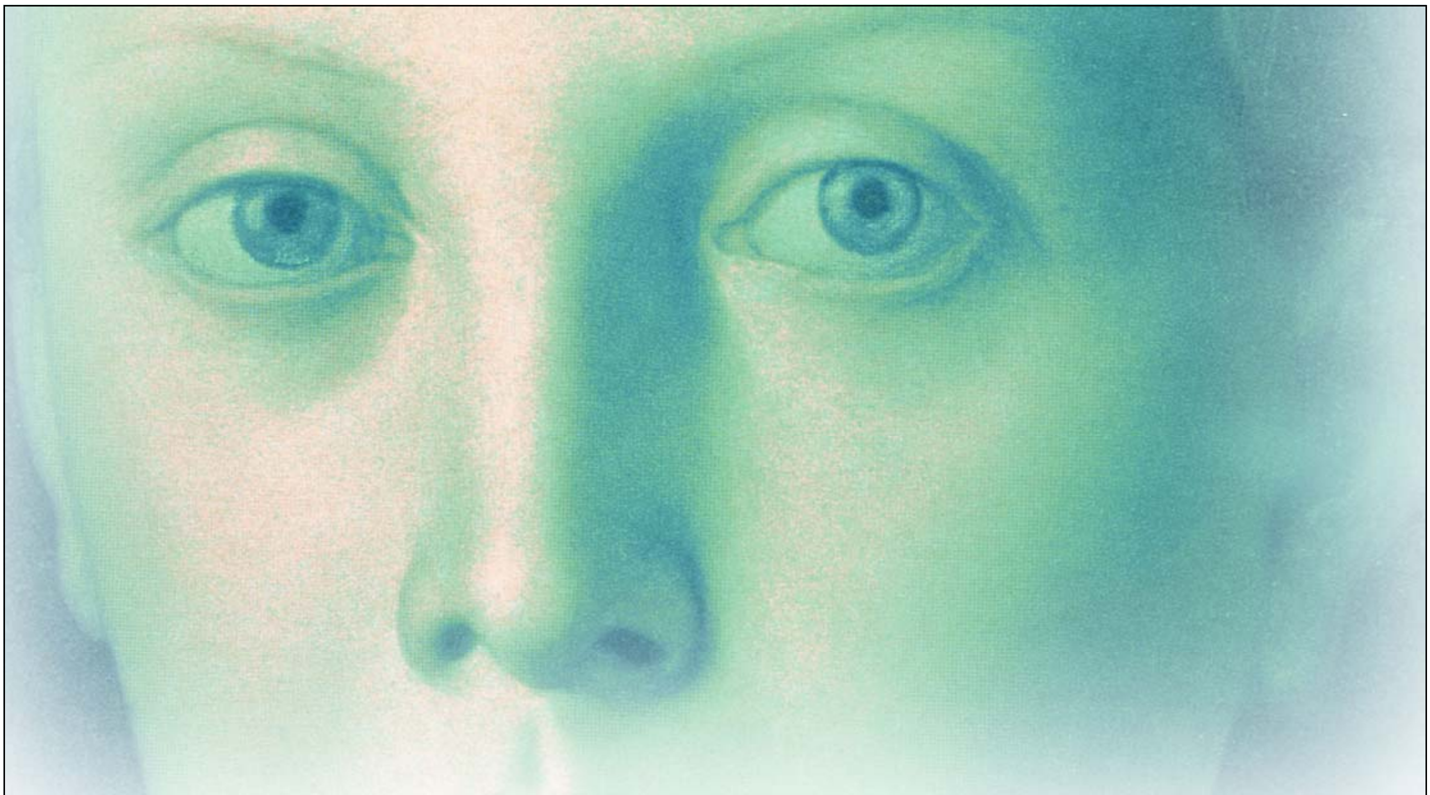


Ipsos-Insight

Actionable Market Segmentation Guaranteed

Part Two of a Two-Part Series on Market Segmentation

A White Paper from the Ipsos Group

Larry Anderson, MS, MBA, Ph.D.
Vice President, Analytics, Ipsos-Insight, Consumer Products Division

Jay L.Weiner, Ph.D.
Senior Vice President, Director of Marketing Sciences, Ipsos-Insight, Services Division

About this issue

Researchers typically focus on the cluster analysis component of market segmentation. However, for segmentation to be successful, one must follow all of the steps of the analytic process, which include:1) clustering, 2) statistically testing the clusters, and 3) profiling the clusters to determine if they satisfy the segmentation objectives. In this paper, *Actionable Market Segmentation Guaranteed – Part Two,* the most frequently proposed clustering procedures will be detailed, as will the methods of statistically evaluating the clusters and profiling the clusters. The pros and cons of various clustering techniques will also be discussed in the context of identifying actionable segments.

About Ipsos-Insight

Ipsos-Insight, the flagship marketing research division of Ipsos in the U.S., has industry specialists serving companies in the following industries: consumer products; technology and communications; health and pharmaceutical; financial services; cable, media and entertainment; agrifood; energy and utilities; and lottery and gaming. Ipsos-Insight provides custom and tracking research services to domestic clients, as well as U.S.-based multinationals. It offers concept and product testing, package testing, attitude and usage studies, omnibuses, tracking systems, brand equity, price optimization and segmentation, marketing models, advanced analytics, and global research. Ipsos-Insight is an Ipsos company, a leading global survey-based market research group.To learn more, please visit **www.ipsos-insight.com.**

Authors

Larry Anderson, MS, MBA, PhD, *Vice President, Analytics, Ipsos-Insight, Consumer Products Division* Jay L.Weiner, PhD, *Senior Vice President, Director of Marketing Sciences, Ipsos-Insight, Services Division*

Executive Editor and Publisher

Amy R. Drill, *Senior Vice President, Product Management & Marketing, Ipsos-Insight,Consumer Products Division*

Research Director and Editor

Lee Markowitz, *Chief Research Officer, Ipsos-Insight, Consumer Products Division*

Design and Illustrations

Barbara Day, *Ipsos North America*

Production

Roland Clifford, *Ipsos North America*

Actionable Market Segmentation – Guaranteed

Part Two of a Two-Part Series on Market Segmentation

Actionable Market Segmentation Guaranteed – Part One demonstrates how the design of the research process is the key to actionable segmentation. Actionable segmentation is a four-step process: (A) determine if segmentation research is appropriate, (B) segment the market using demand variables and profile the segments, (C) evaluate the attractiveness of the segments (define key success criteria) and select targets, and (D) develop the marketing mix for the targets.

Unfortunately, many discussions of segmentation focus only on the analytic process, Steps (B) and (C) of the process. Equally disappointing, most discussions of analytic procedures focus only on cluster analysis techniques (part of Step B) rather than the entire analytic process that includes: 1) clustering, 2) statistically testing the clusters, and 3) profiling the clusters to determine if they satisfy segmentation objectives. In this paper, *Actionable Market Segmentation – PartTwo,* the most frequently proposed clustering procedures will be detailed, as will the methods of statistically evaluating the clusters and profiling the clusters.

Regardless of analytic procedure or sophistication, the analytics cannot guarantee actionable results. Every step of the stepwise approach presented in Part One of this series must be followed to guarantee actionable results.The analytical approach used is important, but it is only one step, and the proposed approach may need to change during the analysis if the original approach fails to produce viable segments. Many multivariate analytical techniques can be used to create post hoc market segments. Because there is no ideal analytic approach that works with every segmentation study, segmentation studies can require the use of two or more approaches to produce the best results.

Introduction

Some researchers confuse cluster analysis with segmentation. Cluster analysis is a mathematical tool used to divide consumers into groups that are similar to each other in some way. Segmentation is a marketing strategy that targets marketing initiatives against subsets of consumers in order to achieve an improved position in the market (higher share, penetration, loyalty, usage, etc.). Therefore, cluster analysis is a tool to help marketers achieve segmentation. For segmentation to be successful, one must follow all the steps with cluster analysis being just a part of the process. The entire segmentation process must be linked to a company's brand or innovation strategy.

This paper will cover best practices across the entire analytic process: 1) cluster analysis, 2) statistically testing the clusters, and 3) profiling the clusters to determine if they satisfy the segmentation objectives. Importantly, the pros and cons of various clustering techniques will also be discussed in this paper in the context of identifying actionable segments.

I. Cluster Analysis

Cluster analysis is a collection of statistical methods used to assign cases to groups or clusters. Cluster analysis methods will always produce a grouping.The groupings produced by cluster analysis may or may not prove useful for classifying objects. If the groupings discriminate between variables not used to do the grouping and those discriminations are useful, then cluster analysis is useful. For example, if grouping zip code areas into fifteen categories based on age, gender, education, and income discriminates between wine drinking behaviors, it would be very useful information if one were interested in expanding a wine store into new areas.

There are two primary forms of clustering tools: hierarchical and non-hierarchical methods. Hierarchical methods build clusters by combining the two most similar respondents or clusters together in a stepwise approach. Non-hierarchical methods attempt to partition the data by splitting the data into subsets. In addition to these two methods, hybrid approaches may also be followed. All three clustering procedures – hierarchical, non-hierarchical and hybrid – will be detailed in this paper.

Issues to Resolve Before Choosing a Clustering Procedure

Before employing a particular clustering technique, one must consider the following: What data should be used to form the clusters? How should similarity be measured and how should clusters be formed? What method should be used to determine the optimum number of clusters? Answers to these questions will eliminate some clustering techniques from consideration. Other techniques will be eliminated because they introduce so much compromise that actionable segments will not emerge. (All methodologies depend on levels of compromise. Perhaps the source of the greatest amount of compromise is in the number of clusters formed. If too few are formed, they might not be differentiated enough to support a targeting effort; if too many are formed, they might not be large enough to be viable targets.)

What data should be used to form the clusters?

The objectives of the research dictate the data that should be collected and used to identify actionable segments, and the data dictates the clustering procedure. For example, if attitudinal segments are desired a series of attitudinal statements could be included in the questionnaire. Attitudinal data is usually captured using scales that are believed to have continuous properties. If a scale is used to collect interval data, like strongly agree to strongly disagree, an appropriate clustering procedure is k-means. On the other hand, if attitude check-off boxes are used to produce frequency data, k-means would be inappropriate. Instead, a hierarchical method or hybrid approach should be employed. Some tools are more appropriate for categorical variables, others for continuous or scale data. Most behavioral information is typically represented by categorical data. Consumers either do or do not purchase/use brands, products, or services.

If a particular analytic procedure is recommended prior to questionnaire design, the questionnaire must be prepared consistent with the data requirements of the clustering methodology. If clustering is requested after the data has been collected, the data will dictate the appropriate clustering techniques. One reason that segmentation studies fail to produce actionable segments is because researchers, armed with a favorite clustering methodology, apply methods to data that require a different approach.

How should similarity be measured and how should clusters be formed?

All cluster methods need a distance measure. Different distance measures may lead to different cluster results. Some distance measures accept only continuous variables like Euclidean distance, and some only categorical variables, such as the simple matching dissimilarity measure used in the k-modes method by Huang (1998). For mixed-type variables, various distance measures exist based on the weighted sum of continuous variables distances and categorical variables distances (Huang 1998, Kaufman and Rousseeuw 1990).You can choose the weight arbitrarily, but improper weight may bias the treatment of different variable types. Banfield and Raftery (1993) introduced a model-based distance measure for data with continuous attributes.They derived this measure from a mixture model, equivalent to the decrease in log-likelihood resulting from merging two clusters. Meila and Heckerman (1998) applied this probabilistic concept and derived another distance measure for data with categorical attributes only.

What method should be used to determine the optimum number of clusters?

Acceptable clustering methods fall into two broad categories: relocation (nonhierarchical) and hierarchical. Relocation clustering methods, like k-means, move cases from one cluster to another, starting from an initial partition, looking for the best fit. The researcher specifies the number of clusters prior to starting the process. Hierarchical clustering methods proceed by stages producing a sequence of partitions in which each one nests into the next partition in the sequence. Unlike k-means, the researcher does not need to establish the number of clusters in advance.

Q-Factor Analysis

The reader of a paper on segmentation analytic methodologies is likely to be familiar with factor analysis, a procedure used to identify clusters of variables that have something in common.Q-factor analysis clusters respondents or cases using the same factor analysis procedures, but by factoring cases rather than variables (the case by variable matrix is transposed prior to factoring) and then observing how cases load on factors to form clusters. Although Q-factor does accomplish data reduction and form clusters, Stewart presented the shortcomings of Q-factor analysis in 1981. Although Q-factor analysis has the strength of reproducibility, many statisticians feel a fundamental discomfort with the assumptions of factor analysis when the data is transposed and correlations are computed between objects and across variables. Although Sharma (2000) suggests that Q-factor analysis not be used to form clusters, some researchers continue to suggest the use of Q-factor analysis.We do not recommend Q-factor analysis as a clustering methodology.

Hierarchical Clustering Procedures

Hierarchical clustering follows one of two approaches: agglomerative and divisive. Agglomerative methods start with each observation as a cluster and with each step combine observations to form clusters until there is only one large cluster. Divisive methods begin with one large cluster and proceed to split into smaller clusters items that are most dissimilar. More formally, a sequence is said to be a hierarchical clustering if there exists two samples, $c1$ and $c2$, which belong in the same cluster at some level k and remain clustered together at all higher levels $> k$.

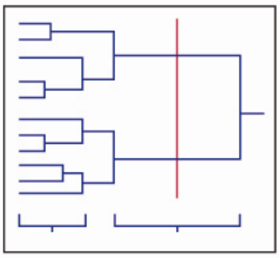
The table below shows various ways of defining inter-cluster distance:

Defining Inter-Cluster Distance	Based on...
Single linkage	The shortest distance between objects
Complete linkage	The longest distance between objects
Average linkage	The average distance between objects
Ward's method	The sum of squares between the two clusters, summed over all variables
Centroid method	The distance between cluster centroids.

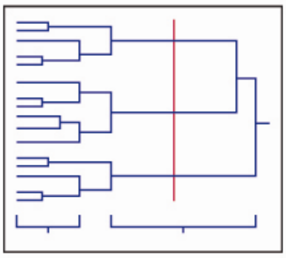
Using Dendograms to Display Hierarchical Clustering Results

A common method of showing the results of a hierarchical clustering is as a dendogram.The first stage dendograms show all samples x_i as singleton clusters. As samples are clustered together using one of the above methods, a final dendogram is formed.

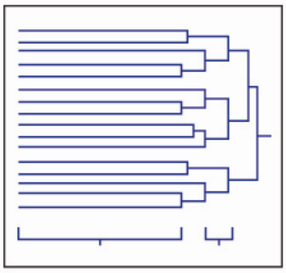
A dendogram that clearly differentiates groups of objects will have short distances in the far branches of the tree (the left part of the diagram below) and long distances in the near branches (the right part of the diagram below). The following is an example of a dendogram that ideally illustrates two clear clusters.



The following dendogram ideally illustrates a clear grouping of three groups.



When the distances on the far branches are long relative to the near branches, then the grouping is not very effective. The following illustrates a dendogram that should be interpreted with great caution.



Non-Hierarchical Clustering Procedures

Non-hierarchical clustering is ***partitioning*** of the sample.

• Each cluster has a seed point and all objects within a prescribed distance are included in that cluster.
• Another way of non-hierarchical clustering is to loop through the sample, assigning each case to the seed point to which it is closest.

Non-hierarchical clustering has three approaches:

1) Sequential threshold...based on assigning all respondents to clusters based on proximity to cluster seeds before the seeds are updated (such as k-means).
2) Parallel threshold...same as above but the seeds are updated as each respondent is assigned.
3) Optimizing...same as the others except it allows for reassignment of objects to another cluster based on some optimizing criterion.

A common approach is to factor analyze variables prior to cluster analysis. This approach can be useful, particularly when some way must be found to reduce a very large number of variables or to prevent a large set of variables measuring a common theme to be over-represented in developing the clusters. However, the Central Limit Theorem assures that when variables are grouped into factors a lot of "smoothing" will occur. Cluster analysis can take advantage of the "lumpiness" of data, and will be impeded by any smoothing that takes place. An alternative approach to using the factors derived from factor analysis directly in cluster analysis is to select one variable most representative of each factor and cluster those variables.

The same argument can be applied to the issue of orthogonalizing variables. That operation would have advantages if it were desirable and appropriate to employ multivariate statistical tests. However, orthogonalizing variables increases the risk of obscuring the very structure to be uncovered.

Standardization and Distance Measures

Usually variables need to be standardized.Variable standardization is needed if the range or scale of one variable is much larger or different from the range of others.

Centering or standardizing across variables within respondents is recommended if resulting segments profile as only high or only low on all variables used to cluster respondents and there is no marketing context for these patterns. Such patterns suggest response bias and it is recommended that response bias be removed.

Ratings from importance or agree/disagree scales may not provide the differentiation required for segmentation. The situation is compounded when there is a large number of ratings and scales have 5 or fewer points. Several approaches have been proposed to increase the level of differentiation. These include:

1) Use 10 point scales, label the extreme points with extreme labels (e.g., Completely or Definitely), ask respondents to look through the entire set of questions before responding, and ask the question within the context of a decision to purchase –"How important is benefit X in your decision to buy one product or another in the category?"
2) Use a constant sum question rotating the attributes – "Assign 10 points across all the attributes based on the importance of each of these benefits in your decision to buy one product or another in the category."

3) Create subsets of benefits using an experimental design and ask respondents to choose the most important and least important benefit in each subset (Cohen & Markowitz, 2002) – this approach is called Maximum Difference Scaling.

Approaches 2 and 3 have the advantage that they are not subject to scale location response patterns (high raters/low raters). Approach 3 has the advantage that it is not subject to scale range response patterns – the tendency to use a small part of the scale or the entire range of the scale. Controlling for these response patterns assumes even more importance in international segmentation where response patterns can vary dramatically among countries.

The statistical distance (Mahalanobis distance) is recommended. This distance measure compensates for intercorrelation among the variables. Often one sums across the within-groups sum-of-products matrices to obtain a pooled covariance matrix for use in statistical distance.

Selecting Seed Points

Let *k* denote the number of clusters to be formed. Fix *k.* (Later you can try *k–1, k+1,* etc.).You choose *k* “seed” points to get started. The results can depend upon the seed points, so often clustering is done several times, starting with different seed points. The *k* initial seeds can arbitrarily be:

• The first *k* cases

• A randomly chosen *k* cases

• *k* specified cases

• Or, chosen from a k-cluster hierarchical solution

The approach of first performing a hierarchical clustering and then performing k-means has lots to recommend. This approach overcomes the weakness of k-means (starting with random points) and the weaknesses of hierarchical clustering (that cases cannot be relocated).

Hybrid Clustering Techniques

Latent Class Segmentation

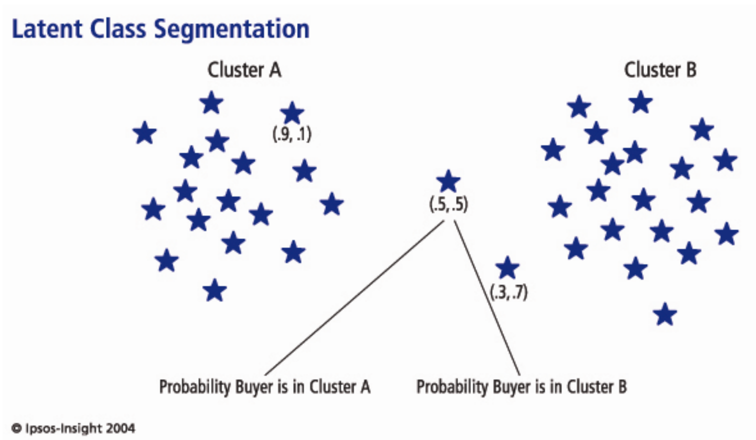
Unlike traditional methods, latent class assigns each respondent a probability of belonging to each cluster. Segment membership can be hard-coded by assigning individuals to the segment with the highest probability. If a respondent exhibits characteristics of more than one segment, they will have a positive probability of falling into each one. Very often in the real world, respondents exhibit the behaviors of more than one segment – they do not fall neatly into one segment. The latent class model captures that ambiguity by using segment probabilities. For example, a given respondent may have a 60% probability of falling in Segment A, a 40% probability of falling in Segment B, and a 0% probability of falling in any other segment.

The Gibbs Sampler, or Markov Chain Monte Carlo Integration Technique,can be used to estimate a latent class model. The Gibbs Sampler relies on random number generators run on high-speed computers. The advantage of using the Gibbs Sampler is that we can take a Bayesian approach to estimate the latent class model. Before the development of high-speed computers, there just wasn’t enough computing power to use Bayesian methods.

In the latent class model, Bayesian methods are conceptually superior to classical methods. Bayesian and classical statistics differ fundamentally in their treatment of probability. Bayesian methods are based on beliefs about the probability of the success of a decision. Conversely, classical statistics are based on the frequency that a decision will be correct. Marketing academics prefer the Bayesian approach because it mimics the way decisions are made in business. That is, a business decision is based on beliefs about the probability of success of that decision rather than the frequency that decision will be correct in an infinite number of trials.

Another advantage to using Bayesian methodologies is that we have the capability to use prior information on segment sizes and behavior for the current period's estimation. This better enables us to identify changes in segment characteristics over time. The Gibbs Sampler provides a means of ensuring that the segments are stable over time, so that changes in segment characteristics are likely to reflect actual changes in the marketplace rather than noise in the data.

In the example below, the outlying respondent floating between the two clusters is equidistant from the two cluster centroids. In traditional segmentation methods, this individual would be assigned to one of the two clusters. The latent class technique, however, describes the particular individual by estimating probabilities as being equally likely to belong to both segments. Latent class segmentation then describes segment populations by weighting the data according to the segment probabilities. LC results can also be converted by "hard-coding" all respondents with a reasonably high-probability leaning toward a specific segment, for more traditional segment interpretation and use.



Latent class segmentation also can be applied to choice and regression models. In this application, latent segments are identified simultaneously with attribute drivers. Within a segment, the relative importance of attribute drivers is the same. Across segments the relative importance of attribute drivers differs.

Latent class segmentation can be employed with Maximum Difference Scaling, the approach that creates subsets of benefits using an experimental design and asks respondents to choose the most important and least important benefit in each subset (Cohen & Markowitz, 2002). This combination approach works well for international segmentation where response patterns can vary dramatically among countries. A latent class analysis approach with Maximum Difference Scaling has the advantage that the segments are derived based on the choices respondents make (which are scale independent) versus cluster analysis of ratings (which are scale dependent).

Two-Step Technique

The type of hybrid clustering we will discuss is based on the SPSS two-step clustering methodology. A key difference between this hybrid approach and others is that the model-based distance measures can be used in situations that include both continuous and categorical variables. In other words, a series of scalar attitudinal questions can be combined with categorical behavioral or demographic variables.

Another advantage of two-step approaches relates to the size of the dataset. Traditional clustering methods are effective and accurate on small datasets, but often fail to scale up to large datasets. Using a two-step approach, these traditional methods will cluster large datasets effectively if these datasets are first reduced into smaller datasets. A sequential approach, which creates clusters and then sub-clusters, is the basic concept of two-step clustering methods. In the first step, a quick sequential cluster method is applied to the large dataset to compress the data and form sub-clusters. In the second step, a cluster method is applied to the sub-clusters to find the desired number of clusters. The records in one sub-cluster should end up in one of the final clusters so the pre-cluster step does not affect the accuracy of the final clustering. Inaccuracy from the pre-cluster step decreases as the number of sub-clusters from the pre-cluster step increases.

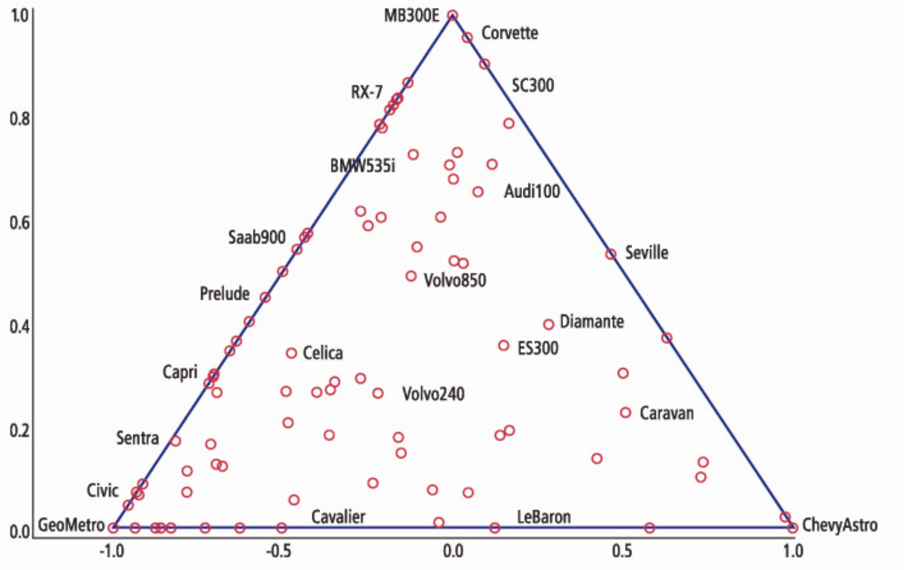
Yet one more – and very important – advantage of the two-step method is that it directly addresses the issue of determining the number of clusters. Other methods apply various strategies to determine the number of clusters. For example, the data might be clustered into a series of numbers of clusters (such as two clusters, three clusters, etc.) and certain criterion statistics calculated for each of them. The one with the best statistic is the "winner." Two-step is based on the Fraley and Raftery (1998) proposal that uses the Bayesian information criterion (BIC) as the criterion statistic or the Banfield and Raftery (1993) approximate weight of evidence (AWE) criterion statistic for their model-based hierarchical clustering. In other words, two-step has a unique approach to determining the optimal number of clusters.

In the end, whatever approach is used, the resulting clusters must satisfy all the criteria for actionable segmentation.

Archetypal Analysis

Archetypal analysis offers a different perspective on the origins and marketing implications of consumer heterogeneity. In many categories, it will be at least as useful to understand "pure types" of customers as to identify multiple "average" customers (as provided by classical segmentation). This procedure should be especially useful to account planners within advertising agencies. The clustering procedures discussed in this paper all have one thing in common: distances from a center point or how close they are to each other. Only a few consumers in a cluster are close to the center of the cluster, or said another way, are represented well by the cluster. Other members of the cluster may be far from the center of the cluster and may be almost as close to another cluster center. For these consumers, a product produced to satisfy the expressed needs of the cluster will necessarily be a compromise. In fact, to some degree, producing a product for a cluster results in a compromise for every cluster member. Archetypal analysis eliminates the compromise.

Archetypal analysis also provides rich visual data representation. The maps produced by archetypal analysis visually show associations and do not require language interpretation or common references. Archetypal analysis supports concept development in that, for example, the ideal vehicle for a Baby Boomer is a combination of the archetypes in the following figure:



Source: Riedesel, Paul (2002) "Archetypal Analysis in Marketing Research: A New Way of Understanding Consumer Heterogeneity," ARTF, 2002

The Archetypal algorithm will use the data from a survey (or other data set) to construct archetypes and, per the figure, the MB300E, ChevyAstro, and GeoMetro. The algorithm will then describe each targeted segment vehicle as a mixture of these archetypes. Most researchers agree that it is much easier (and effective) to evaluate how the mixture of attributes (cars in this example) changes over time than to review how specific attributes/lifestyles change. The mathematical problem is to locate archetypes that optimally describe the entire dataset taking into account the ideal attributes of vehicles across all targets.

In summary, archetypal provides the following benefits:

1) The procedure recognizes that data points rarely fall into mutually exclusive clusters.

2) The procedure outputs visual maps that are easily interpretable.

3) The procedure shows degrees of similarity/dissimilarity between different data points.

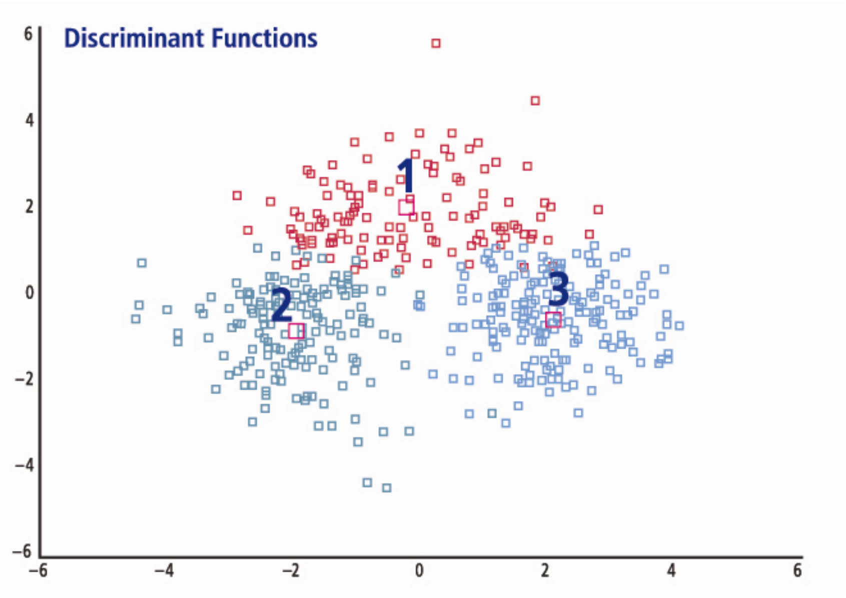
4) Information about the similarities/dissimilarities between data points can be used in other analyses.

II. Methods to Statistically Test Clusters

Discriminant Analysis

Cluster analysis may be used in conjunction with discriminant function analysis. After multivariate data are collected, observations are grouped using cluster analysis. Discriminant function analysis is then used on the resulting groups to discover the linear structure of either the measures used in the cluster analysis and/or different measures.

The following chart of the members of a three-cluster solution suggests that the members of the three clusters are quite different from each other. If it is found that the clusters meet all the criteria for an actionable segmentation, they will be further described and become named segments.



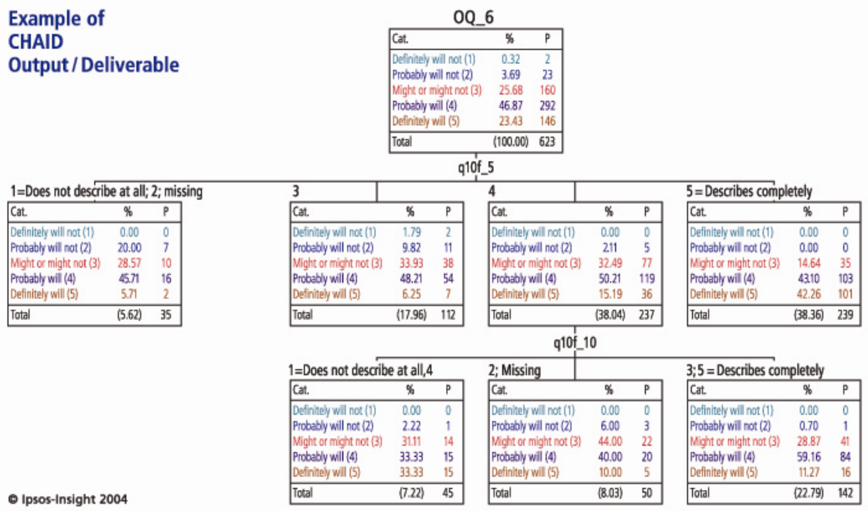
If, for example, it is found that cluster #1 represents a lifestyle, or attitudes towards purchase of a product consistent with the goals of the research, it is possible that a concentrated effort could be made to target this segment. On the other hand, if all three segments should be approached but with different messages, like which features of a product need to be promoted, this conclusion will be clear from the discriminant deliverables.

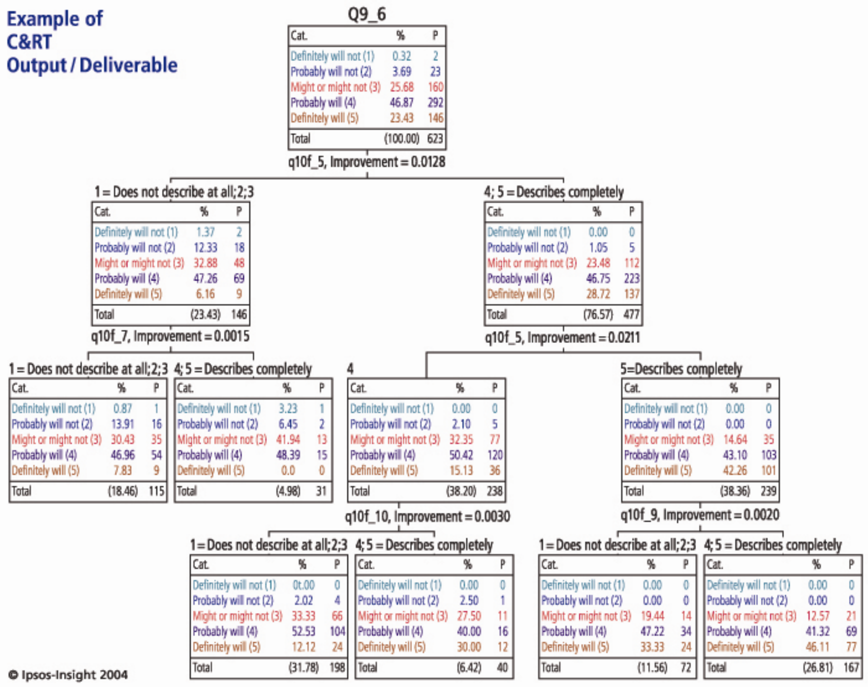
Classification Algorithms

Discriminant analysis will be used to classify respondents into the pre-defined clusters based on variables used or not used to create the clusters. The classification or scoring algorithm can be used to identify potential members of each segment for future research. The "hit rate" or proper classification of cluster members offers a sense of the quality of the clustering.

CHAID/C&RT

CHAID (**Ch**i-square **A**utomatic **I**nteraction **D**etection) is a statistical technique used to determine how a set of independent variables best predicts the dependent variable. CHAID works with both categorical and continuous level measures. CHAID determines how best to split each independent variable. The resulting model is a classification tree that shows how the key independent measures predict the dependent variable. C&RT (Classification and Regression Trees) is similar to CHAID, but the splits are binary and the rules for splitting are different.





The initial box of a CHAID or C&RT output includes all respondents in the study. The next level of the tree provides the best variables and the split in this variable that provides the best discrimination of the dependent measure. For CHAID this process continues until there is no variable that provides incremental discrimination (or the number of respondents in the box is too small to split). For C&RT, the splitting occurs well beyond what is needed for final segments. Then the tree is pruned based on a cross-validation procedure.

For example, our dependent measure might be level of satisfaction (measured on a five point scale) and our independent measures might be various ratings of product/ service delivery (also measured on five point scales). The attribute "overall satisfaction with account representative" might provide the best predictor of overall satisfaction.

Applications of CHAID and C&RT

CHAID and C&RT are used for a variety of marketing research applications. Both CHAID and C&RT are useful techniques for developing a scoring algorithm to predict a segmentation solution. These scoring trees are easily programmed into CATI recruiting programs for small base qualitative studies (including focus groups). CHAID and C&RT can also be used for customer (or employee) satisfaction studies to help determine the drivers of satisfaction. Finally, both CHAID and C&RT have been used extensively in data mining. Financial institutions can use decision trees to predict credit worthiness. Direct marketing companies can use decision trees to predict acceptance of an offer.

Random Forests

Leo Breiman (2004) has developed an extension of tree-based clustering algorithms (CHAID/CART) called Random Forests. The main advantages of Random Forests are: 1) significant increase in the accuracy of prediction, 2) trees are not subject to over-fitting, and 3) many input variables can be used and processed quickly. The primary disadvantage is that Random Forests cannot be shown visually so they can appear to be a black box, like a neural net.

Random Forests represent the collection of many tree-based clusters, typically 100 or more. Each tree is based on a random selection of respondents of size *n* drawn *with replacement* from the total sample of size *n.* On average about two-thirds of the original sample is selected for each tree, the "training" sample. The remaining sample is called the out-of-the-bag (OOB) sample. Now, assume there are *m* predictors available for splitting the sample. As each tree is grown, a random subset (*k*) of the total number of predictors (*m*) is made available for splitting at each node (*k,* the number of predictors available, is constant for every node in every tree although the actual *"k"* predictors are selected at random for each split). The final nodes in each decision tree provide a classification for each respondent in each training sample. Next, the OOB respondents in each decision tree are classified by taking these respondents through the decision trees. For each respondent for each decision tree, a "vote" is cast based upon which class the respondent falls in the final nodes of the tree. Then, for each respondent, the total "votes" are counted across all decision trees and the respondent is assigned to the class with the most votes.

III. Profiling Clusters to Determine if Segmentation

Objectives are Satisfied

Of all the complex analytic tasks involved in segmentation research, profiling of clusters is by far the easiest. By examining the properties of the cluster centroids it is possible to develop a robust description of the cluster. For example, the means of attribute statements for each cluster centroid will provide the first indication of whether or not the segmentation objectives are being accomplished.

Simple cross-tabs with a cluster banner can be used to identify cluster differences on demographic, lifestyle, lifestage, geographic, behavioral, attitudinal, or other variables used to target. Statistical procedures can be used to improve the efficiency of this process, but it is as important as any other stage of the research. Profiling makes the clusters come alive and become segments.

Tracking Segment Growth

The accuracy of tracking a segment's growth may not get a lot of attention because many studies fail to produce actionable targets. Why track a segment that is not an actionable target? If the focus is placed on identifying actionable segments, not homogeneous subsets that are the figment of a researcher's imagination, tracking will be easier and algorithms, like Fisher's coefficients, will perform reliably.

Regardless, tracking will remain a challenging segmentation task. Dynamic segmentation provides an immense amount of valuable information in categories in which consumer behavior or product offerings change frequently. This type of segmentation provides all the behavioral, demographic, attitudinal, and psychographic detail of a standard behavioral segmentation study, plus how those variables change over time. Latent class techniques lend themselves to dynamic segmentation due to the ability to use prior probability information to produce the cluster solutions in subsequent waves.

Summary

Actionable segmentation requires a sound research process based on four steps: (A) understanding the business/strategy issues to determine if segmentation is appropriate, (B) employing an analytic approach that creates and profiles differentiated segments, (C) evaluating the attractiveness of the segments by profiling on key success criteria and selecting targets, and (D) developing an efficient marketing plan. As part of this four-step process, a complete analytic process is employed: 1) selecting an appropriate clustering approach, 2) statistically testing the clusters, and 3) profiling the clusters to determine if they satisfy segmentation objectives. Employing a complete analytic process is necessary but not sufficient to guarantee actionable segmentation.

There is no analytical approach that is universally appropriate – the approach must be selected based on the segmentation objectives and information collected. Next, the analytical approach is evaluated based on its ability to produce differentiated segments that profile differently on the key success criteria and can be efficiently targeted. If the analytical approach does not meet these objectives (segments that can be differentiated on key success criteria and can be efficiently targeted), then additional analytic approaches are evaluated until these objectives are met.

Summary of Methods

Segmentation Objective	Data Transformation	Clustering Methodology
General understanding of a market, products/brands used	Binomial/frequency	• Hierarchical • Latent class
Benefits sought…find segments that allow for differentiation of products/services to make them more meaningful to customers; identify ideal products	Scalar	• K-Means • Two-step
	Binomial/frequency	• Hierarchical • Latent class • Two-step
Identify attitudes of targets and segments that exhibit attitudinal profiles that are consistent with marketing efforts	Scalar – K-Means	• K-Means • Two-step
Profile product purchase and use patterns like heavy users, or frequent flyers	Binomial/frequency	• Classify • Cross-tab • Latent class
Improve competitive positioning – Fine-tune current marketing strategies; gauge the company's market position (how the company is perceived by its customers and potential customers relative to the competition)	Scales only – K-Means combination	Two-step
Product use/Attribute preference (check-off)	Binomial/frequency	• Hierarchical • Latent class
Brand loyalty (loyalty/customer value segments)	Binomial /frequency	Proprietary algorithm
Brand equity (core loyals, prospects, vulnerables, price shoppers, system beaters)	Binomial/frequency	Proprietary algorithm
Shape marketing mix (products, promotion, distribution and pricing) to fit the most promising markets or to rule out changes to the mix	Scales only – K-Means combination	Two-step
Price sensitivity – Deal Proneness – Price sensitivity by purchase/use patterns	Binomial/frequency	• Hierarchical • Latent class
	Promotional preference	Two-step
Advertising…media use, AIO, hybrid studies	Usage – frequency	• Hierarchical • Latent class
Distribution…store loyalty and patronage, benefits sought	Usage – frequency	• Hierarchical • Latent class
Realize economies by concentrating on profitable products or services	Cluster products	Hierarchical
Realize economies by targeting marketing and selling efforts	Cluster customers	Depends on data type
Simplify marketing effort. It is easier to address the needs of smaller groups of customers, particularly if they have many characteristics in common (e.g., seek the same benefits, same age, gender, etc.)	Multiple sources of data	Two-step
Identify opportunities (niches)…Identify under-served or un-served markets. Using "niche marketing," segmentation can allow a new company or new product to target less contested buyers and help a mature product seek new buyers	Binomial/frequency	• Hierarchical • Latent class
Score databases by assigning market segmentation characteristics to database records and generate prioritized prospecting lists	Scalar – concern about reproducibility	CCA K-Means
Maximize Lifetime Value of the product by repositioning during each stage of the product life cycle (PLC)	Binomial/frequency	Profile consumer classes who buy at each stage of PLC
Measure category involvement level	Scalar	Proprietary algorithm
Identify segments that maximize attractiveness measures	Combine results of proprietary algorithms (categorical data), demographics (latent class), scalar questions	Two-step

About the Authors

Larry Anderson joined Ipsos in 1997 bringing decades of research experience having been employed by the IRS, USAF, several Fortune 500 companies, and five universities. His list of publications includes articles on segmentation, perceptual mapping, consumer behavior, customer satisfaction, nonparametric forecasting models, and GIS. During the past few years he has presented papers on Syndicated Segmentation, Data Considerations When Mapping Breast Cancer Incidence, Brand Equity Measurement, Public Utility Customer Claim Verification, DTC Advertising, Rx Compliance, Brand Equity, and The Influence of Children on Purchasing. Although Larry grew up in Texas and was awarded his Ph.D. by the University of Texas, he now works in New York and calls Vermont home. You can reach him at **larry.anderson@ipsos-na.com.**

Jay L.Weiner has his Ph.D. in Marketing/Research from the University of Texas at Arlington, as well as an M.B.A. in Finance and Commercial Banking and a B.B.A. in Marketing. Jay has published and presented papers on Conjoint, Choice and Pricing Research in conference proceedings. At Ipsos, Jay consults with many leading corporations on marketing, and market research issues. He assists them in using advanced methods to make better marketing and business decisions both domestically and globally. Jay has expertise in Pricing, Segmentation, Loyalty, Customer/Employee Satisfaction, Conjoint Analysis and Discrete Choice Analysis in addition to solid multivariate statistical skills. You can reach him at **jay.weiner@ipsos-na.com.**

References

Banfield, J. D., and A. E. Raftery. 1993. Model-based Gaussian and non-Gaussian clustering. *Biometrics* 49:803–821.

Bowman, D., Heilman C.M., and D.B. Seetharaman. 2004. Determinants of Product Use Compliance Behavior. *Journal of Marketing Research* 41:3.

Breiman, L. 2004. Random Forests. (In press, Machine Learning).

Cohen, S.H., and P. Markowitz. 2002. Renewing Market Segmentation: Some new tools to correct old problems. *ESOMAR 2002 Conference Proceedings,* 595–612. Amsterdam: ESOMAR.

Fraley, C. 1998. Algorithms for model-based Gaussian hierarchical clustering. *SIAM Journal on Scientific Computing* 20: 270–281.

Fraley, C., and A.E. Raftery. 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. *Computer Journal* 93:294–302.

Gibson, Lawrence D. 2001. Is Something Rotten in Segmentation? *Marketing Research* 13 (1): 20–26.

Green, Paul, and Donald Tull. 1988. *Research for Marketing Decisions.* Englewood Cliffs, New Jersey: Prentice Hall.

Hair, Joseph F., Rolph E. Anderson, Ronald L.Tatham, and William C. Black. 1998. *Multivariate Data Analysis.* Upper Saddle River, New Jersey: Prentice Hall.

Huang, Z. 1998. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery* 2 (3): 283–304.

Kaufman, L., and P.J. Rousseeuw. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis.* New York:Wiley.

Meila, M., and D. Heckerman. 1998. An experimental comparison of several clustering and initialization methods. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence,* 386–395. San Francisco: Morgan Kaufmann.

Myers, James H. 1996. *Segmentation and Positioning for Strategic Marketing Decisions.* Chicago: American Marketing Association.

Riedesel, Paul. Archetypal Analysis in Marketing Research: A New Way of Understanding Consumer Heterogeneity. http://www.action-research.com/archtype.html.

Sambandam, Rajan. 2003. Cluster Analysis Gets Complicated. *Marketing Research* 15 (1): 16.

Sharma, Subhash. 1996. *Applied Multivariate Techniques.* New York:Wiley.

Theodoridis, S., and K. Koutroumbas. 1999. *Pattern Recognition.* New York: Academic Press.

Wedel, Michel, and Wagner A. Kamakura. 1998. *Market Segmentation: Conceptual and methodological foundations.* Boston: Kluwer Academic.

Weinstein, Art. 1994. *Market Segmentation: Using demographics, psychographics, and other niche marketing techniques to predict and model customer behavior.* Chicago: Probus Pub. Co.

Wind,Yoram. 1978. Issues and Advances in Segmentation Research. *Journal of Marketing Research* 15 (000003):317–339.

Wind,Yoram and Susan P. Douglas. 1972. International Market Segmentation. *European Journal of Marketing* 6 (1): 17–25.

Wittink, Dick R., and Philippe Cattin. 1989. Commercial Use of Conjoint Analysis: An Update. *Journal of Marketing* 53 (3): 91-96.

Zhang,Tian, Raghu Ramakrishnan, and Miron Livny. 1996. BIRCH: An efficient data clustering method for very large databases. *Proceedings of the ACM SIGMOD Conference on Management of Data,* 103–114. New York: ACM Press.