QUIRK'S
Marketing Research Review

SURVEY OF ANALYSIS METHODS

Part II: Segmentation Analysis

By: Rajan Sambandam

Editor's note: Rajan Sambandam is vice president/ research at The Response Center, a Fort Washington, Pa., research firm. He can be reached at 215-641-2251 or at rsambandam@response-center.com

Segmentation analysis has been a part of marketing research for decades. It continues to be useful in a variety of different situations, even when the primary objective of the study is not segmentation. Since segmentation divides the data into comparatively homogenous groups, marketing efforts such as targeting, positioning, retention and product development can be more efficiently performed. While the value of segmentation analysis is rarely questioned, the methods of developing segments have always given rise to considerable debate.

One of the simplest ways of segmenting the data is basic cross tabulation analysis. Respondents can be divided into, say, age or income groups and their differences studied across a variety of questions. This approach of pre-defining the respondent is often referred to as *a priori* segmentation

Use of a priori segments, while attractive, is often not sufficient given the need to obtain complex segments based on multiple variables. Therefore, most of the time segments need to be developed after data have been collected. In this article we will consider various segmentation methods, both traditional and recent, that can be used to address marketing research problems. The three methods we will consider are:

- Cluster Analysis

- Neural Networks

- Mixture Models

It should be noted at the outset that regardless of the method used for analysis, the quality of the segmentation scheme is determined by its usefulness to the manager. Even if the statistics indicate that a particular solution is the best one, if it is not useful to the manager then the segmentation analysis should be seen as a failure. This condition is not as harsh as it seems, because not only are many different solutions possible with a given set of variables, but changing the variable set can lead to more solutions. Further, using different analytical methods can also provide new solutions. Finally, there is also the option of dividing some of the segments obtained into sub-segments, if that would make them more actionable.

Next, we will look at each of the segmentation methods mentioned above and how they work. This will be followed by a discussion on ideas for developing good segments.

CLUSTER ANALYSIS

Cluster analysis is the traditional method used for segmentation in marketing research. This is actually a family of methods that subsumes many variations and can be broadly classified under two distinct groups: *hierarchical* and *non-hierarchical* (or partitioning) methods.

Hierarchical clustering includes methods where the basic idea is to start with each observation as one cluster. Each observation is located on an n-dimensional space where n is the number of attributes used in the analysis. The distances between observations are measured using some form of distance metric such as Euclidean distance. Based on these distances, observations that are closest to one another are joined together to form a new cluster. This process continues until all observations have been merged into a single cluster. The optimal number of clusters can be determined by looking at standard measures of fit (statistics such as the cubic clustering criterion, pseudo-f and pseudo-t^2) provided for each cluster solution.

Conversely, it is possible to start with all observations together as one cluster and work backwards until each observation becomes a cluster by itself. With both variants of the hierarchical method, the analyst will have to study the results of the analysis to determine the appropriate number of clusters.

In the non-hierarchical methods (such as k-means clustering), random observations are chosen as seeds (or cluster centers) for a pre-specified number of clusters. Thus, the initial ordering of the data can dictate the formation of clusters. Observations that are closest to a particular seed are assigned to that seed, thus giving rise to clusters. The analyst then obtains the fit statistics for a variety of solutions in order to determine the optimal number of clusters.

Choosing the appropriate number of clusters is never easy even with data sets that are reasonably well behaved. In commonly used methods like k-means clustering, the analyst needs to specify the number of clusters desired. This can be problematic, because the algorithm will assign observations to clusters regardless of whether there are bona fide segments in the data. The fit statistics that indicate the optimal number of clusters are often unclear. Sometimes the optimal number of clusters may not make operational sense. In such cases actionability should be considered before deciding on the optimal number of clusters. Hence, the process of developing segments from data using cluster analysis has a high interpretive content.

NEURAL NETWORKS

Artificial neural networks are a recent addition to the variety of techniques used for data analysis. There are two basic types of neural networks: *supervised learning* and *unsupervised learning networks*. Supervised learning networks can be used in place of traditional methods like regression and discriminant analysis and were discussed in the previous article in this series. Unsupervised learning networks are the subject of our discussion here.

Unsupervised learning networks are generally used when there are no clear distinctions between dependent and independent variables in the data and when pattern or structure recognition is required. Since pattern recognition is really what is needed in segmentation analysis, unsupervised neural networks can be used for this purpose. The type of unsupervised learning network most appropriate for the problem of segmentation is the Self-Organizing Map (SOM) developed by Teuvo Kohonen.

Self-Organizing Map

A typical SOM consists of an *input layer* and a grid like structure known as the *Kohonen layer*. The input layer contains the variables that are going to be used in the analysis, while the Kohonen layer is a grid of processing elements. Each of the variables in the input layer is connected to each of the processing elements in the Kohonen layer. These connections have random starting weights attached to them before the start of the analysis.

When the information from the first respondent is presented to the network, the processing elements "compete" with each other. By mathematically combining the first respondent's score on each input variable with the weight of each connection, the processing element with the "winning" score can be determined. Winning implies that this particular processing element is the one that most closely resembles the input scores of the respondent. This processing element is called the "winner". The weights associated with the winner will then be adjusted to more closely resemble the respondent. The network can be thought of as learning the response pattern of the respondent.

Not only are the weights associated with the winning processing element changed, but the weights of the neighboring processing elements are also changed. In other words an area of the grid is learning the response tendencies of the respondent.

When the second respondent's data are presented to the network the process is repeated. If the second respondent is similar to the first, then a processing element from the same area of the grid wins. Whether it is the same processing element as the last time will depend on whether the second respondent is exactly similar to the first one. If the second respondent is very different, then a processing element in a different part of the network will win.

At the end of this process the grid will show a two dimensional representation of the data with different segments showing up as different neighborhoods on the map. Because of the iterative process described above, substantial segments cannot be formed around outliers. This is a clear advantage this method enjoys over traditional k-means cluster analysis.

SOMs also have an advantage in that they were initially developed as not just a data-reduction tool, but also as a data visualization tool. This capability allows the SOM to provide a more intuitive understanding of the relationship between the variables and the segments, hence making the process of developing segments easier. However, some experts feel the reduction of a multidimensional problem to a two-dimensional space for visualization can actually be a disadvantage because of the constraints it may impose on the segmenting process. A further disadvantage in the case of large datasets is the amount of time required to run the analysis as compared to k-means cluster analysis.

MIXTURE MODELS

This is another broad category of segmentation methods. The basic idea linking methods in this category is that the data contain many distributions or segments which are mixed together. The task of the analysis then becomes one of unmixing the distributions and for this reason they are also called unmixing models.

One of the major differences between the cluster methods described previously and mixture models is the prior specification of the number of segments in the data. In non-hierarchical cluster analysis we have to explicitly specify the number of clusters in the data. In hierarchical cluster analysis the results are presented for every possible cluster solution (with the limit being each observation treated as a cluster), thus effectively making the analyst choose the optimal number of clusters. In mixture models, the assumption of underlying distributions allows the use of optimization approaches that can automatically identify the number of segments (distributions) in the data.

Another variation of the mixture model approach to segmentation is known *as latent segmentation analysis*. While it belongs to the mixture model family, it has some advantages that might be very useful in a marketing research context. For example, latent segmentation analysis makes it possible to simultaneously conduct a segmentation and key driver analysis, where each segment can have its own unique key driver analysis. Thus if a manager is interested in not just identifying segments but also understanding the key drivers of, say, satisfaction within each segment, this would be an appropriate method to use. This process is more efficient than running a segmentation analysis first, followed by separate key driver runs for each segment.

While mixture models can be very useful in creating segments, they also have some disadvantages. The primary disadvantage is with the large amount of time required to run the analysis, especially when compared to k-means cluster analysis. There are also other disadvantages such as sensitivity to the presence of outliers.

CONCLUSION

While different types of approaches to segmentation analysis have been discussed here it is not clear that there is one approach that is the best in every situation. Segmentation analysis often involves trying more than one method to obtain the best result. The main reason for this is that unlike key driver analysis, segmentation analysis is quite unstructured. The final solution depends on the number and nature of variables included in the analysis. Changing even one variable can have a strong impact on the results. Without seeing the results, however, it is hard to identify the variables that can be useful in the analysis. This type of circular problem implies that the most important step in a segmentation analysis is the choice of variables to use. The more thought we put into selecting the variables, the more likely it is that the results will useful.
There are a few other steps that can be taken (with any of the methods described here) to increase the chances of developing good segments. These are:

- Eliminating outliers

- Using as few input variables as possible

- Using input variables with low correlation between them

Eliminating outliers not only ensures that segments don’t center on them, they also result in tighter, better-defined segments. Using as few input variables as possible is hard to do, but very important for deriving useful and timely solutions. Beyond the fact that irrelevant variables can sabotage the analysis, using too many variables complicates the analysis leading to solutions that are not useful. One way of reducing the number of input variables is to remove those that are highly correlated with other input variables. Further, since segmentation methods don’t work as well when there is a collinearity problem in the input variable set, it makes sense to eliminate collinearity as much as possible.