MARKET SEGMENTATION

The simplest and most effective way to operate an organization is to deliver one product or service that meets the needs of one type of customer. However, to the delight of many organizations over the years, broadening one's product or service portfolio by appealing to new customers and/or expanding the portfolio's success with current customers can lead to profitable business growth.

Consequently, most organizations are on the lookout for new products and services, or variations on current ones, that meet different needs or wants of current customers or which bring new customers into the fold. Market segmentation is a set of concepts and tools that can guide management thinking and lead to new profitable product or service offerings.

When a company moves beyond one product or service offered to one type of customer, it has started down the path of segmentation. If the product traditionally appeals to men, perhaps a variation will appeal specifically to women, thus segmenting the customer base. If the product appeals to young adults, perhaps a variation will appeal to 'tweeners. There may be a product that appeals to 'tweener males vs. 'tweener females.

What began as segmentation based solely on gender can be extended to segmentation based on many, many factors. The key is to find a clear-cut homogenous group of customers or potential customers whose demand for a uniquely configured product or service supports a sustainable market opportunity.

The rationale for segmenting is that customers within segments typically have more in common with each other than with customers in the remaining segments. This commonality helps focus marketing efforts toward each segment. Basically, customers who share similar characteristics on key defining attributes are presumed to have similar needs, and this similarity allows development of specialized products and/or messages that uniquely addresses those needs.

The Basics of Market Segmentation

A successful market segmentation initiative answers the following critical business questions:
* How can we a

Market segmentation guides strategic and tactical decisions around:

*

Recommended Strategies with Acquisition and Retention		Customer Status	
		Actual Customer	Non-Customer
Predicted Customer Status	Predicted Customer	Thanks	Acquisition

	Predicted Non-Customer	**Retention**	**Lowest Priority**

Recommended Strategies with Existing Customers		**Purchase Volume**	
		High Volume	**Low Volume**
Predicted Purchase Volume	**Predicted High Volume**	**High Touch**	**Up-sell & Cross-Sell**
	Predicted Low Volume	**Recognition**	**Virtual Channel**

However, new products and services come with risks. There are tangible costs such as development, production, distribution, marketing, pricing and sales, along with intangible costs such lost opportunities or diminished brand appeal.

Segmenting customers and prospects into groups with shared characteristics is a frequently used method of limiting risk. In effect this approach allows a company to move toward a "one product for one customer" business model.

Market segmentation is a broad term that covers a variety of approaches to analyzing customers. Broadly speaking, it focuses on assigning customers to groups that can be further analyzed for a number of different issues.

There are many different ways that the general research goal of segmentation can be achieved. Typically, these revolve around two key decisions that must be made in the process of developing segments:

- How are the variable(s) used to create the segmentation determined?
- Once the variables have been determined, what statistical methods will be used to determine the number of segments present in the sample data.

Unfortunately, there are no clear-cut “correct” answers to either of these questions, but there are ramifications that ripple out from the decisions made at each stage. Segmentation methods — more so than many other areas of statistical analysis — combine art and interpretation to a great degree.

Determining Variables

Basically, the decision with regard to selecting the variables that will be used in the segmentation ultimately comes down to two different choices:

- Variables can be determined by **management perception of business opportunities,** because of the "world view" that managers have of their business world. In its simplest form, this sometimes leads to segmentation models that are based on a single criterion, such as which product a customer purchases, or the region of the country where a customer or prospect resides. Alternatively, complex multivariate models can be developed with this approach where project managers select measures based on an *a priori* view of the factors perceived as most relevant in differentiating customers and prospects.

- Determining the variables to use in a segmentation model can also be done on the basis of **statistical criteria**. For example, statistical models predicting sales volume may be carried out, and the key drivers in these models used to select the measures that are used in a subsequent segmentation model.

Many different statistical techniques, as well as outcome measures, may be used to determine the variables that are ultimately used in the segmentation analysis. If this approach is used, those decisions are typically based upon the data collected and other overall project objectives.

Of course, it is also possible to use **hybrid models** where some of the measures used in the segmentation are drawn from the statistical approach, and others are based on management decisions regarding what they want to see included in the model.

Is there one "best" approach? Simply put, there is not. The management guidance approach is potentially problematic because it explicitly ignores what the market is saying, and as such can lead the analysis in a direction that may not be terribly fruitful. On the other hand, blind reliance on statistical modeling procedures can be inherently conservative. There is a tendency to focus on past trends and to allow them to drive the analysis to a great degree. If the market is undergoing (or about to undergo) a revolutionary change, then the statistical modeling approach may miss opportunities that an insightful management may anticipate.

Regardless, for each of these approaches the 1st selection of variables need not necessarily be the last. In particular, subsequent analysis may well indicate — in fact, it usually does — that some of the measures selected for use in the segmentation model really add little value and simply confound and confuse the model.

Statistical Methods

There are two broad classes of statistical models[1] that are used in segmentation models, and within each of these classes there are numerous specific algorithms that yield different results. The two primary statistical approaches to developing segmentation models include:

[1] In addition to the two classes discussed in this paper, some consider the hybrid predictive models of Automatic Interaction Detection (AID) and Chi-Square Automatic Interaction Detection (CHAID) as approaches to segmentation that focus on splitting the sample into segments as opposed to joining respondents into segments.

- **Hierarchical Methods** develop segments from the ground up. A matrix is calculated that determines how similar each respondent in the data file is to every other respondent. The most similar respondents are joined together, the similarity matrix is re-calculated, and the next-most-similar respondents are joined. This process continues until all respondents are joined together.

  At each stage, diagnostic data are presented that help to determine if this stage constitutes a good stopping point in the segmentation process.

  While this sounds quite simple in theory, there are over 20 different statistical algorithms — 11 of which are quite commonly available — that are used in the hierarchical segmentation models to achieve this general goal.

  One of the key benefits of the hierarchical models is the fact that assorted diagnostic data are provided to help inform decisions regarding the proper number of segments to retain for the segmentation model, but even with these diagnostics there are no clear-cut unambiguous rules. A key limitation of all of the hierarchical segmentation models — more true of some of the methods than others — is that they can produce segments which in multidimensional space resemble elongated chains as opposed to the typically desired tight, spheroidal clusters.

- **Iterative Centroid Methods** develop in a more deductive fashion. Given a set of variables *and* a specification of the number of segments to retain, a set of initial starting points or centroids equal to the number of desired segments is defined for each measure. Respondents are then added to the segment that they are most similar to, the centroids are re-calculated, and respondents are re-assigned to the new centroid that best characterizes their pattern of scores on the measures. This process keeps repeating (iterating) until no respondents change segments, and the centroids stop drifting.

There are 7 major variations in the iterative centroid methods – three of which control the vast majority of uses – available for the analysis.

The strength of this approach is that it tends to produce the tight, spheroidal segments that are typically preferred for subsequent analysis. The primary limitation is that it requires *a priori* determination of the number of segments, and provides few diagnostics that help to determine the number of segments that should be used in the analysis.

What Do We Recommend?

Our usual approach to developing segment models includes the following stages:

- Work with the client to select a key outcome criterion or criteria, and develop multivariate models that identify the key drivers that characterize the data.

- Review the key driver data with the client and suggest a set of measures for the segmentation analysis, augmenting this recommendation with other measures as desired by the client.

- Review the data on the proposed measures, and select 2-3 different hierarchical clustering algorithms that appear best suited for the project and data at hand.

- For each of the 2-3 hierarchical clustering algorithms, perform a Monte Carlo probability simulation study with a minimum of 500 simulation "runs" to determine the most likely number of segments that would be suggested by this analysis. Typically, this will suggest a small range that should be tested.

- If necessary, select 2-3 of the different iterative centroid methods and test alternative models representing each of the different solutions for the number of clusters, as determined by the previously completed Monte Carlo probability study. Note that often by this stage it is not necessary to test alternative algorithms since the data will usually be quite clear as to the best approach to use.

- Review the findings of the iterative centroid models and, if necessary refine by dropping measures that are clearly random and do not contribute to the segmentation structure.

Client involvement at each stage of the process is a welcome addition that can help to ensure management buy-in for the final segmentation model.

Hierarchical Clustering Methods

Method	Comments
Average Linkage	Unweighted pair group method of clustering using arithmetic averages
Centroid Method	Unweighted pair-group method using centroids
Complete Linkage	Bases the segmentation on minimizing the distance of the most distant member of the segment
Density Linkage	Nonparametric probability density model that has three major classes of estimates for defining "density:"
	a) Density estimates are computed from a preliminary cluster analysis using the k-means method
	b) Density is defined by specification of a number of "neighbors" to use in calculation.
	c) Define a sphere are each data point, with the number of respondents within that sphere being used to calculate the density function.
Equal Variance Maximum Likelihood Method	A segmentation model that relaxes the rigid requirement that a respondent be assigned to a single cluster and allows estimation of the probability of segment membership.
Exact Optimization	A segmentation model that uses both the similarity among the measures in the data set but also the sequential positioning of the records within the data set, i.e., sequential or "close" records are favored. Often used in files where records are sorted by time and sequence "makes sense."
Flexible Beta	The Lance-Williams flexible beta method. The beta parameter allows you to check for different cluster shapes in multidimensional space.
McQuitty Similarity	Weighted pair-group method which minimizes the average linkages among members in the segment.
Minimum Mutation	A biological model that focuses on minimizing the mutations required when moving from a simpler to more complex model.
Median	A weighted pair-group method using centroids defined by medians as opposed to averages.
Single Linkage	Simplest form of segmentation where the nearest neighbor can define a join. Tends to produce cluster chains.
Triads Algoritm	A variation on the complete linkage model (see above) that extends the usual pair-wise calculation of distance into looking at triads.
Two-Stage Density Linkage	A variation of density linkage model where all individual records must be assigned to a modal cluster before modal clusters are allowed to join together.
	a) Density estimates are computed from a preliminary cluster analysis using the k-means method
	b) Density is defined by specification of a number of "neighbors" to use in calculation.
	c) Define a sphere around each data point, with the number of respondents within that sphere being used to calculate the density function.
Two-Way Joining	Another segmentation model that uses both the similarity among the measures in the data set but also the sequential positioning of the records within the data set. See also Exact Optimization methods.
Ward's Minimum Variance	Minimize the within-group error sum-of-squares. Tends to produce equal sized clusters.

Centroid Clustering Methods

Minimization Function	Assignment Process		
	Non-Iterative	Iterative with Drifting Centroids	Iterative with Static Centroids
Mean Absolute Deviation from Median			
Root mean *p*th-power absolute deviation			
Maximum Absolute Deviation			

Sensitivity to Outliers

Sensitivity to Initial Seed Points for Algorithm