Archetypal Analysis in Marketing Research:
A New Way of Understanding Consumer Heterogeneity

Dr. Paul Riedesel
President, Action Marketing Research
paulr@action-research.com   612-879-9212

Introduction

Archetypal analysis is a method for analyzing and understanding the heterogeneity of consumers (or products) in a market. The technique is well established in the physical science and mathematics, but was only introduced to the field of marketing research in 1998. This paper compares archetypal analysis to more traditional market segmentation based on cluster analysis.

Consumer Heterogeneity

Consumers (a term meant to include business buyers) differ systematically in their attitudes, needs, behavior, and other characteristics. Because it is usually impractical for a firm to create a unique marketing mix for each consumer, firms rely on various means of grouping customers. Mass marketing is one extreme: everyone is offered the same product(s), the same advertising and promotion, the same distribution channels, the same prices. More sophisticated marketers employ *market segmentation* where consumers are divided into two or more classes with a different marketing mix for each class. The rationale for market segmentation is well-established and can be found in any marketing or consumer behavior textbook. It holds that consumers will more readily respond (buy!) when offered products, messages, prices, etc. that are closer to their needs.

Markets may be segmented on any number of possible bases, but brand-oriented marketers and especially advertising agencies have long preferred *attitudinal segmentation*. It traditionally uses large-scale survey research in which representative consumers respond to a battery of attitude questions. Various multivariate techniques, most often non-hierarchical clustering, are used to define a limited number of segments and to assign each respondent to one of them. (There are exceptions, but this is the rule). Using segmentation research, a marketer can:

- Estimate the size of each segment and its share of category volume
- Select one or more segments to target
- Create a marketing mix tailored to each target segment

Unfortunately, no matter how good the segmentation analysis, the classification of consumers into a few discrete segments is always arbitrary to some degree. Experienced researchers know this, but the underlying fictional nature of discrete segments is quickly forgotten. When cleverly named segments take on a life of their

own, researchers are gratified to see their work actually making a difference. The data typically used for segmentation (attitudinal or otherwise) consist of measures on m variables for each of n respondents. The analyst is faced with an n x m "cloud" of data points; usually using cluster analysis, the analyst's job is to divide that multivariate cloud into a small number of sections or segments. If a division can be made that is useful, then the segmentation was a success. Still, it behooves us to remember that the segments are rarely tidy boxes.

Consider an example with just two continuous variables.
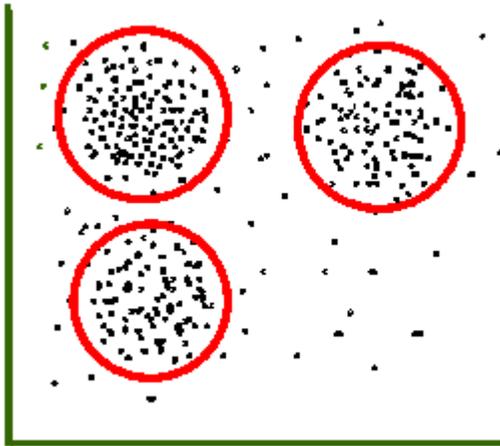
Each dot is a consumer. There are three obvious segments here, so those within a certain distance of the mid-point of each segment are assigned to that segment (the red circles). It is common but not universal practice to leave outlying cases unclassified. That is, consumers who are outside the three circles are not considered to belong to any segment. Even though this example shows an exceptionally clear segmentation, the fact remains that there are far more than three kinds of consumers. And there are quite a few unclassified outliers. We could reduce the number of outliers by making the segments more inclusive, but that also make them more heterogenous.

This type of classification is not "wrong." It is often very useful. But it is only one way of understanding consumer heterogeneity.

Archetypal Analysis: An Alternative Paradigm

Archetypal analysis is an alternative to traditional segmentation for understanding consumer heterogeneity.

- Segmentation (at least when based on statistical cluster analysis) makes the implicit assumption that there are several "average" consumers. They are found in the dead center, statistically speaking, of each cluster.
- Archetypal analysis assumes instead that there are several "pure" consumers who are on the "edges" of the data. All others are considered to be mixtures of these pure types.

Archetypal imagery is common in everyday language.

- "There's a little bit of _____ in everyone."
- "After he/she was born, they threw away the mold."
- "There's only one . . .
    - Martha Stewart
    - John Wayne
    - Maria Callas
- The Real McCoy

It is not their extremeness that makes certain consumers archtypical. Rather, it is the fact that a given set of archetypes in a survey optimally explain the variation among other *other* respondents. As explained below, archetypes are identified through objective statistical procedures. Yet in our experience, archtypical consumers are more intuitive, more familiar than the types of consumers which seem to represent traditional cluster-based segments.

When consumers encounter advertising directed at archetypes rather than semi-fictional segment averages, they will be more likely to recognize the objects of the advertising. It is true that most consumers are mixtures of several archetypes, but many will be dominated by a given archetype (a statistic which can be calculated) and in any case the type of consumer evoked in such advertising will ring true (advertising verité?)

Application of Archetypal Analysis

Since the procedure is quite new to marketing research, its applications are not fully developed. At the very least, archetypal analysis is appropriate:

- In categories where the archetypes are likely to be *aspirational* in nature.
    - Whether a category is aspirational is a matter of judgment, not something that archeypal analysis can tell you
    - The goal would be to aim messages at what consumers *want to be*, not what they *are*.
- In categories where the extreme types of consumer are the trend-setters
    - Speak to them and the others will follow

Research Procedure

The basic mathematics for doing archetypal analysis are most accessible in a paper by Professors Adele Cutler and Leo Breiman called "Archetypal Analysis" in the journal *Technometrics*, November, 1994, Volume 36, Number 4, pp. 338-347.

In non-technical terms, the analysis uses a matrix of data with n respondents and m variables--just like most segmentation analysis. The algorithm solves several regression equations at once with the object of maximally explaining the variance within the data. The user must specify the number of archetypes to be solved for. If/when a solution is found, we know:

- The archetype weights for each respondent
    - They sum to 1.00 for each respondent

- o If a person has a weight of 1.00 on an archetype, he or she *is* the archetype
- The relative dominance of each archetype in the sample
  - o This is just the mean of all respondents' weights on that archetype
- The degree to which each variable is associated with each archetype

Illustration

Assume we have collected seven psychographic measures having to do with lawn care and chose to define three archetypes. We may observe the following data values for the three archetypes.

	Archetype 1	Archetype 2	Archetype 3
Status-Seeking	7	2	4
Comfort-Seeking	3	5	7
Economical	2	7	4
Achievement Orientation	6	1	4
Aesthetic Orientation	6	2	3
Control Seeking	7	4	2
Functional Orientation	2	7	5
MEAN MIXTURE WEIGHT	0.24	0.63	0.13

The first archetype is instantly recognizable as someone who for whom a nice lawn is a matter of pride and who will spare no expense to achieve it. The second archetype is recognizable as someone who is practical about lawn care--just get the job done. According to the mean mixture weights, this archetype is dominant in the sample. Finally the third archetype is someone who is unusually concerned about making lawn care easy. Although these data are fictional, the archetypes we describe are familiar to anyone in the industry.

The other key output from archetypal analysis in marketing research is a set of weightings or mixtures for each respondent. Following are the mixture weights for seven people on the three archetypes we just described.

	Archetype 1	Archetype 2	Archetype 3
Dave	1.00	0.00	0.00
Leonard	0.00	1.00	0.00
Jean	0.00	0.00	1.00
Julie	0.47	0.27	0.26
Roger	0.00	0.81	0.19
Megan	0.27	0.08	0.65
Chris	0.38	0.32	0.30

Note that Dave *is* the first archetypal consumer, Leonard *is* the second archetype, and Jean *is* the third type. That is, they have a mixture weight of 1.00 on one of the archetypes. Julie is more like Dave (since her largest weight is on archetype 1), Roger is more like Leonard, and Megan is more like Jean. Finally, Chris is an undifferentiated mixture of the three archetypes.

Further Analysis

Once an archetypal solution has been agreed on, many other types of data analysis are possible.

- You will certainly want to know the mean mixture weights, which show the relative dominance of each archetype
- You will want to examine the distribution of those weights--are they very skewed?
- You may wish to classify respondents into groups which are dominated by one archetype or another
    - The choice of the cutoff is arbitrary (as in segmentation!)
    - You may have a large group of respondents who are undifferentiated, that is, are not dominated by any archetype
    - This is an intuitively realistic situation; even with a powerful segmentation, it is often hard to pigeonhole many consumers into a single segment
- These groupings (probably including an undifferentiated group) can be compared on other variables in the survey, e.g.
    - purchase interest
    - purchase behavior, product usage
    - other attitudinal data
- It can be insightful to compare archetypal data to segments you have otherwise defined
- The mixture weights could be used as linear predictors in modeling or could be correlated with other continuous variables
- Given comparable data from fresh respondents, they can be assigned mixture weights
    - Identifying a reduced set of variables that optimally reproduce the mixture weights remains a statistical challenge, however

Summary

Archetypal analysis offers an entirely new 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.