Data Driven Segmentation

Chapter 10

Rajiv Grover, Dean

Fogelman College of Business and Economics

University of Memphis

Memphis, TN 38152

DATA DRIVEN SEGMENTATION

Perhaps nothing is more integral to the challenge of developing a sound marketing strategy than the process of market segmentation. As explained in Chapters 1 and 3 of this text, market segmentation involves the marketing manager finding the best matches between the firm's most distinctive competences in general and product characteristics in particular, and those specific customers who most value them. A firm's key competences typically express themselves in the particular sets of features or attributes associated with their products and services. For example, the very popular and successful iPod music and video player is a clear expression of what many feel is the distinctive core competence of Apple Computer – its remarkable innovativeness. This product is stylish, contemporary, cool, unique, and extremely functional. Such product attributes are interpreted by customers as the unique benefits of purchasing and using Apple's offerings. Exploiting its distinctive competence, a firm chooses a target segment for its product by identifying the subsets (segments) of customers whose specific needs and wants are best addressed by the product. Furthermore, the firm develops a marketing mix to profitably satisfy the targeted group of customers' needs and wants. A firm that markets a diversified portfolio of products does so for each and every product in its portfolio while keeping the synergistic issues discussed in Chapter 1 in mind.

Chapter 3 examined the relationship between a product's value proposition, and segmenting and targeting strategies. As we clearly saw in that chapter, considerable creativity can be required in order to identify potential segments and decide which ones to target. Consider the example of TiVo – the well-known digital video recorder brand. TiVo allows users to record and save television shows, skip commercials, and also provides advanced viewing features such as pausing, rewinding and watching in slow motion. It also suggests programs to the viewer and automatically tracks their broadcast to save them if required. But what segments actually make up the market for digital video recorders? Creative thinking on the part of a marketer could identify at least three different segments to potentially target. One segment could consist of people who, due to their busy schedules, are often faced with the problem of missing their favorite television shows. They could be anybody from middle to upper level business managers, to moms with young children. These are individuals who have very hectic work and personal schedules, and who also have very strong preferences for viewing certain television programs. TiVo's ability to record and save missed shows, as well as its ability to compress viewing time through the elimination of commercials, offers substantial benefits to these types of busy individuals. Another target segment to consider could simply be heavy television users. They love watching television and often face the dilemma of wanting to watch two different shows that air at exactly the same time on different channels. TiVo's record and save features can allow these individuals to maximize their television viewing. TiVo's recommendation and preference engine could also be very useful to these heavy television users in identifying programs to consider watching. A third segment could be identified based on TiVo's ability to function as a "super" VCR. Individuals desiring to download their videos and photos easily onto the TiVo DVR, as well as those who want to move the contents of their old video tapes onto the TiVo device, could constitute a sizeable segment. Other segments can be identified based on its pause and slo mo features. The point that was emphasized in Chapter 3 was that in some cases depending upon the segment(s) targeted, there might be a need to add other features to the offering to fully satisfy the needs of the segment or some features may be deleted to save the segment some costs. For example, if TiVo were to be positioned as a super VCR, ease of copying pictures and videos from cameras and old cassettes to the disk would be a high-importance feature. So also would be feature of being able to replace the disk when it was full.

> Point to Ponder: How would you price a TiVo positioned as a super-VCR? Would you pitch the subscription service to the segment desiring TiVo for this purpose? What other features would you propose for the heavy-users and busy segments? Are there other possible segments, e.g., professional sports or sports enthusiasts that TiVo can leverage? Should it have a line of products, at least one for each segment?

When TiVo was a relatively new product, these segments would not very likely have been identified based on examining the numerical findings from customer responses to market research surveys. Instead, they would be the result of understanding the marketplace, combined with creative thinking on the part of a marketing manager. Such creative analysis for segmentation is entirely suitable for some types of marketing problems – problems where an innovative product is introduced to satisfy certain unique needs that are not yet being satisfied with existing products. Data-driven segmentation is quite different. Here, the data themselves are used to reveal the existence of important market segments. Generally speaking, such types of data are available for products that are well understood by the market. In these cases, data add a layer of concreteness to creativity in terms of the size of the market and other finer preferences and intentions. Consider the following example. A computer manufacturer wants to know what people feel and think about computers and related technologies. Specifically, the firm wants to know what people feel about their knowledge of and enthusiasm toward computers and their attitudes in general toward technology – whether it's good or bad for society, etc. The firm also wants to know what kinds of segments exist as well as how substantial they are. The implications for marketing mix decisions are significant. If the firm finds that there is a substantial group of people who have positive attitudes toward computers, but who are not very knowledgeable and are apprehensive about their ability to understand computers better – an educational campaign could be launched to help these individuals gain the requisite knowledge in order to become more viable customers. Similarly, the firm could organize and implement a targeted public relations campaign if it learned that a substantial segment of relatively influential people feel that computers are bad for society. The firm could collect useful data from a random sample of consumers on a relevant set of variables (see Table 10.1), and segment the market based on how the sample responds.

INSERT TABLE 10.1 HERE

Any given individual will respond very similarly to some in the sample of respondents, and at the same time very differently than others. The key objective of data-driven market segmentation is to use actual data to identify groupings of customers where the members of a given group are most similar to other members within that group (homogeneity within segments), but are measurably different (dissimilar) from other customers who are members of other groups (heterogeneity between segments). These measures of similarity/dissimilarity can be based on any of a number of relevant geographic, demographic, psychographic, behavioral, or benefits-related characteristics of the overall market being segmented. These represent some of the potential *bases* of segmentation.

Segments can be defined in terms of single or multiple variables (bases). And they may be defined a priori, or they may be empirically determined after data analysis. A priori segmentation implies that the variables on which the market is to be segmented and their specific values (i.e., cut-offs) are chosen or determined upfront because of strategic reasons. Empirically determined segments uncover the important variables and their cut-offs (e.g., age range, income levels, etc.) through statistical analysis. An example of using a single variable, determined *a priori*, might be to segment a particular market based on age. Using the age when many individuals start raising families to distinguish between the segments, two different segments could be identified – adults 18 to 29 years old, and adults 30 and above (i.e., cutoff equals 30 years old). Likewise, the firm could identify three different segments based on age, and offer different products for children, young adults, and older adults. Another example of *a priori* segmentation based on a single variable is the business-to-business firm that segments its customers based on their size (e.g., number of employees) – small, medium, and large. In these cases, the only measured variable used to determine an individual's (company's) membership in a given segment is their age (size), and the cut-offs were determined a priori. Clearly, if age is the single basis for defining the segments, then each segment (at least with respect to age) will be homogeneous within (members of the same segment will have similar ages) and heterogeneous between (members of different segments will differ with respect to their ages). Since younger individuals typically desire different product features and benefits than older individuals, the firm may decide to target each segment with a different product, or to target only one of the segments. Ideally, each segment will also have other homogeneous characteristics, such as common media consumption habits, that make reaching them easier. In this instance, age is the *basis* of segmentation while other characteristics such as media habits are used to *profile* the segments.

Alternatively, the firm may decide on an a priori basis to segment the market based on two different variables – age and income. If the firm decides to represent an individual's income as either high or low, and a person's age as either young or old, four potential segments are identified – younger with high income, younger with low income, older with high income, and older with low income. A firm like General Motors, with the resources and capabilities to offer many different products to many different market segments, may decide to position and target its offerings using these two variables. The Hummer brand may be targeted to younger individuals with relatively high incomes. The Chevy Cobalt could appeal to younger adults with relatively low incomes. Cadillac would belong in the choice set of older adults with higher incomes. And the Saturn brand could be positioned to address the needs of older individuals with lower incomes. Of course, more than two different variables can be used as the *bases* of segmentation. Then each segment may be further *profiled* using other variables. Personality traits, lifestyle characteristics, gender, and media preferences could be used by General Motors to develop more precise segment profiles. It's important to realize that any given customer characteristic (e.g., age, income, personality, etc) could potentially be used either as a basis of segmentation, or as a means to profile the segment. Whether the variable (characteristic) is used as a basis for segmenting, or for profiling the segments, is often a function of its ability to influence the homogeneity within and the heterogeneity between segments. Variables that most significantly affect the homogeneity within and heterogeneity between segments should generally be the first variables considered as potential bases of segmentation. However, the reasons for segmentation might be the over-riding factor. For example, a firm might want to segment the market based on price sensitivity in order to launch a sales promotion campaign.

Which variables to potentially use to segment the market can be chosen a priori, based on the creativity and intuition of the marketer. This creative component is critical, since it motivates and influences exactly what types of customer information (demographic characteristics, attitudes, lifestyles, etc.) must be collected and examined. However, using creativity to judge which of the chosen variables will be significant and what are their cut-offs (e.g., the age below which they are classified into one group, and above which they are classified into another group), may not always be easy or even possible. Referring back to our computer example, it may not be possible to intuitively examine the different measured variables and determine what are the most significant variables as well as the specific levels on which segments substantially differ. In such cases, detailed analysis of the actual data may be necessary for segmentation. Here, statistical techniques such as cluster analysis, factor analysis plus cluster analysis, and latent class regression, among others, come into play. The data that are used for empirically segmenting markets can be collected via survey research (attitudes, lifestyles, psychographics) or they may be collected as actual behavioral data (e.g., purchases, brand, quantity, time, frequency, place, price, media consumed, etc.). Regardless of the data source, the idea is to understand the structure of the overall market and to develop appropriate target market and marketing mix strategies. Thus, data-driven segmentation may be used to understand the price-sensitive customers (and different variants of them) in order to target them with specific promotional efforts, or to group potential customers together based on their attitudes and lifestyles. This data-driven aspect of market segmentation is the focus of this chapter. We will examine several important analytical techniques for identifying homogeneous market segments and developing precise market segment profiles.

> Point to Ponder: When would a-priori segmentation suffice? When would just a description of the targeted segment suffice (e.g., the company will target youngsters who use their mobile phones extensively and generally do not care much about their phone bills) without the need for segment sizes and measures of segments' characteristics?

CLUSTER ANALYSIS

One of the most common and popular methods for analyzing data in order to define and understand market segments is cluster analysis. Cluster analysis is a mathematical method for classifying individuals or objects into groups or "clusters" or segments on the basis of their similarities. This methodology supports the objective of identifying groupings of customers who are most similar to other customers within a given segment, but who are distinct from other customers who are members of other segments. The objective of cluster analysis is to identify a reasonable number of market segments that are as homogeneous as possible with respect to a set of important characteristics which form the basis of segmentation. Two types of analytical methods for determining segments using cluster analysis are hierarchical and non-hierarchical clustering.

Hierarchical Clustering

To introduce how Hierarchical Clustering works, let us consider the example of a beverage company that is planning to introduce its new fruit drink, and is struggling to determine which potential consumers are the best prospects to initially target.[i] The firm has developed a product that it believes has two key attributes – it tastes good and it's a healthy beverage. There are several important questions to answer. Will any consumers like the taste? Will any consumers believe that the drink is nutritious? Will any consumers think that the beverage is both tasty and healthy? If so, how many consumers are likely to perceive the product as a viable combination of these generally mutually exclusive attributes? If there are not enough consumers who view the product as a unique combination of taste and health, will the firm have to abandon the idea altogether, since being perceived as only tasty or only healthy may not offer the firm any significant competitive advantage? With these two variables forming the bases of segmentation, analysis of data collected from a sample of fruit drink users, using one of a variety of statistical techniques will be required to answer these questions. Table 10.2 presents hypothetical data for a set of customers, indicating their taste response to the new beverage on a scale from 1- 50, and their nutrition rating on a scale from 1- 5. Figure 10.1 locates each of these customers as intersecting points in the two-dimensional space defined by these two variables (taste and nutrition).

INSERT TABLE 10.2 HERE

INSERT FIGURE 10.1 HERE

It is obvious from Table 10.2 that some individuals respond very positively to the taste of the new fruit drink, while others do not. Likewise, some individuals perceive it to be a much more nutritious beverage than others. A visual examination of Figure 10.1 suggests that there appear to be four groups, or clusters, based on measures of taste and nutrition. Said differently, there appears to be four groups of individuals who are in closer proximity (i.e., more similar) to each other than they are to individuals in any of the other three groups. The measure of similarity used to visually identify the four groups is simply the straight line distance between each possible pair of consumers. Unfortunately, visual clustering is not feasible in most actual marketing applications that involve significantly larger sample sizes and many more than two customer variables. In such instances, statistical techniques are required to make sense of the data and hierarchical clustering is an analytical method that offers a solution.

A Simple Analytical Solution

Using our fruit drink example, the straight line distance between any two customers in the dimensional space defined by their taste response and nutrition rating is used to measure their similarity. More specifically, the distance between any two consumers in our example can be calculated as the sum of the squared distances between those two customers for each of the customer characteristics being considered. For example, based on the data in Table 10.2 the squared distance between CUS1 and CUS2 would be the squared difference in their taste responses plus the squared difference in their nutrition rating, or:

$$(42 - 44)^2 + (5 - 4)^2 = 5$$

A summary of the squared distances between each possible pair of customers, referred to as a dissimilarity matrix, can be easily constructed. For our fruit beverage data, this matrix is provided below in Table 10.3.

INSERT TABLE 10.3 HERE

> Point to Ponder: What do you think it means to use Euclidean distance (sum of the squared distances)? Can you think of other ways to measure the distance between two objects? Could clusters change if you used a different distance measure?

Small values in this matrix indicate relative similarity between paired customers, while larger values indicate relative dissimilarity. It is obvious from Table 10.3 that CUS2 and CUS3 enjoy the lowest degree of dissimilarity (i.e., highest similarity) based on taste and nutrition, with a squared distance of 4.25. It would be reasonable to infer that these two customers, more so than many other pairs of customers, could be members of the same potential market segment. Likewise, CUS3 and CUS7 have the highest degree of dissimilarity (squared distance equals

1156.25), indicating that these sets of customers, based on these particular segmenting characteristics, do not likely belong to the same segment.

> Point to Ponder: How would you extend the above formula for distance measure between two individuals if there were three variables on which they responded, i.e., they also responded to the question asking for their perceptions on the product's shelf-life? What would be formula if more variables were added?

Hierarchical cluster analysis assumes that the marketing manager has no prior knowledge or view of exactly how many clusters optimally describe the data. The analysis starts from the premise that each individual customer is a cluster unto itself. The process then involves successive clustering iterations where individuals (and/or clusters of individuals) are grouped together based on their squared distance from each other, until only one cluster comprised of all individuals remains. Starting with an 12 cluster solution (each individual is his/her own cluster), we systematically continue adding individuals to clusters as we work our way down to a one cluster solution where every individual is a common member of one overall cluster.

> Point to Ponder: Do you think it would be possible to develop a process where all individuals start off in one cluster and they are then successively broken up into more and more groups? How would you decide how to make the first split in the group?

What's the Distance from One Cluster to Another?

Let's assume that we start with the premise that CUS2 and CUS3, based on their low dissimilarity (high similarity), belong to the same segment. Based on merging these two customers into one cluster, we are now left with 11 total clusters (CUS2 and CUS3, plus each of the remaining ten individual customers). An important question becomes, how should we now represent the new cluster consisting of CUS2 and CUS3 in a new dissimilarity matrix now made up of 11 (instead of 12) members? In other words, how should we measure the distance between two clusters when at least one of the clusters has multiple members? In our example, the specific question becomes what value should be used to indicate the taste response and nutrition rating for the cluster consisting of CUS2 and CUS3, in order to determine its distance from the other nine remaining customers? Different options are available to answer for this question. The most common approaches used include the Centroid, Single Linkage (Nearest Neighbor), Complete Linkage (Farthest Neighbor), and the Average Linkage methods. Figure 10.2 graphically displays the distance between clusters using each of these methods.

INSERT FIGURE 10.2 HERE

Centroid Method. Using the Centroid method, the average value (centroid) of the characteristics of all cluster members is used to represent the cluster. In other words, the cluster is assumed to consist of one average member, whose characteristics are represented as the average values of all cluster members. For example, the cluster comprised of CUS2 and CUS3 is assumed to have the characteristics of a hypothetical average member whose taste reaction is 45 ([44 + 46] / 2) and whose nutrition rating is 4.25 ([4 + 4.5] / 2). This new cluster, with its average taste and nutrition values, is now named and used to create a new dissimilarity matrix (see Table 10.4 and Table 10.5). The process continues by identifying the next pair of consumers (or clusters) who are most similar to each other (i.e., lowest dissimilarity score), and so on, until only one overall cluster remains.

Single Linkage (Nearest Neighbor) Method. This method defines the similarity, or distance, between any two clusters as the minimum distance between all possible pairs of individuals comprising the clusters. For example, the distance between our cluster consisting of CUS2 and CUS3, and the cluster consisting only of CUS1, would be calculated as the *minimum* of the distance between CUS1 and CUS2, and CUS1 and CUS3. According to Table 10.3, the distance between CUS1 and CUS2 is 5, while the distance between CUS1 and CUS3 is 16.25. As a result, 5 would be used to represent the distance between these two clusters.

Complete Linkage (Farthest Neighbor) Method. With the complete linkage method, the similarity between any two clusters is calculated as the maximum distance between all possible pairs of individuals comprising the clusters. For example, the distance between our cluster consisting of CUS2 and CUS3, and the cluster consisting only of CUS1, would be calculated as the *maximum* of the distance between CUS1 and CUS2, and CUS1 and CUS3. According to Table 10.3, the distance between CUS1 and CUS2 is 5, while the distance between CUS1 and CUS3 is 16.25. As a result, 16.25 would be used to represent the distance between these two clusters.

Average Linkage Method. This method defines the similarity, or distance, between any two clusters as the average of the minimum and maximum distances. For example, the distance between our cluster consisting of CUS2 and CUS3, and the cluster consisting only of CUS1, would be calculated as the *average* of the distance between CUS1 and CUS2, and CUS1 and CUS3. According to Table 10.3, the distance between CUS1 and CUS2 is 5, while the distance between CUS1 and CUS3 is 16.25. As a result, 10.625 ([5 + 16.25] / 2) would be used to represent the distance between these two clusters.

Point to Ponder: How do you think these different linking methods will change the nature of the resulting clusters? When do you think a manager might prefer one of these linking methods over another linking method?

As mentioned above, using any of these approaches, a new cluster with its associated taste and nutrition values is now identified and used to create a new dissimilarity matrix. Tables 10.4 and 10.5 describe an eleven cluster solution using the centroid method to value the new cluster (CLUS 11). This process would continue by identifying the next pair of consumers (or clusters) who are most similar to each other (i.e., lowest dissimilarity score), and so on, until only one overall cluster remains.

INSERT TABLE 10.4 HERE

INSERT TABLE 10.5 HERE

Different Methods, Different Solutions

The above hierarchical clustering methods differ slightly with respect to how the similarity, or distance, between clusters is calculated. However, each method may perform differently based on the quality and orientation of the data collected and used in the analysis. Studies designed to assess the relative performance of these various methods have identified certain effects that should be carefully considered in any segmentation analysis. One such effect is a *chaining effect*, where one particular cluster simply keeps growing as it adds members with successive clustering iterations to eventually form one very large cluster. Single linkage, because it focuses on the minimum distance between clusters pairs is typically more susceptible to chaining effects than other hierarchical clustering methods. In many marketing applications, chaining is not desirable as it tends to result in a single overall cluster or segment, and may disguise the identity of other more useful and operational segments.

In addition, these methods will respond differently to data that is contaminated with *outliers* that are situated at significant distances from other data points. Specifically, when there are outliers in the data, single linkage and complete linkage may produce very different results. While single linkage will tend to combine clusters due to the existence of outliers that are close together, complete linkage will not. Complete linkage tends to identify more homogeneous, compact segments. Due to the difficulty in determining the optimal clustering method for any given set of data, it is often appropriate to use and compare the results from *all* of these various methods in order to determine the optimal approach. However, finding significantly different cluster solutions using these different methods can be a warning sign that no natural clusters exist within the data.

> Point to Ponder: If there were no outliers and there were a few distinct clusters inherent in the data, i.e., large inter-group differences and low intra-group differences, would you say that all approaches would give very similar results?

How Many Clusters?

For our twelve customer example, hierarchical clustering methods would compute eleven different cluster solutions; including the one cluster solution that would include all twelve customers (see Table 10.6). The dendogram in Figure 10.3 translates the data from Table 10.6

into a visual representation of that information. The numbered boxes at the bottom of the Figure correspond to the 12 customers in this example. At the lowest level of the dendogram, each customer is a cluster. As you go up the dendogram, the first linkage you encounter is between CUS2 and CUS3, indicating that those two customers are closest and should be jointed to form the first segment. Proceeding up the dendogram we see that the second segment joins CUS8 and CUS9 and that the third joins CUS4 and CUS6. The next level of the dendogram joins the CUS2-CUS4 segment with the CUS4-CUS6 segment, etc. At the top of the dendogram all customers are joined into a single segment.

INSERT TABLE 10.6 HERE

INSERT FIGURE 10.3 HERE

The relevant question becomes what is the optimal number of clusters? In general, a good cluster solution is one that has within-cluster homogeneity and between-cluster heterogeneity. In other words, in order to provide useful insights for potential marketing mix decisions, the customers within each cluster should be as similar to each other as possible with respect to their ratings of the new drink's taste and nutritional value, and at the same time the different clusters should be relatively distinct or different from each other. Various measures for assessing the optimality of cluster solutions are provided by most popular software packages such as **SAS** and **SPSS**. All measures are based on some combination and usage of within-group homogeneity and between-group heterogeneity. In general, measures of within-cluster homogeneity and between-cluster heterogeneity can be thought of as similar to measures of variance. For within-cluster homogeneity, the variance between the various members within a cluster can be calculated. Low values suggest greater within-cluster homogeneity. For between-cluster heterogeneity, the variance between the cluster centroids can be calculated. If combining two clusters significantly raises the within-cluster variance, this is an indication that two relatively heterogeneous clusters have been combined.

Although it is beyond the scope of this text to examine the actual measures and mathematical procedures, the student nonetheless requires some basic understanding of what are small or large values of homogeneity/heterogeneity. Small and large values are assessed in a relative fashion. Thus if there is an unusually large increase in the total amount of within group variance on combining two clusters, one might argue that these two clusters should not really be combined. Hence, further aggregation of clusters is not warranted. Figure 10.4 shows the plot of the degree of within segment heterogeneity for the beverage company example. Notice the unusually large increase in within-cluster distance that occurs when we go from the 2 cluster solution to the 1 cluster solution. (Analysts use the term "elbow" to refer to the kink in the plot between the two cluster and the one cluster solutions.) Hence, it appears that the two cluster solution is optimal.

INSERT FIGURE 10.4 HERE

> Point to Ponder: Why did our visual inspection of the data suggest the possibility of a four cluster solution, while the analysis of the data provided in Figure 10.4 indicates a two cluster solution? Is it because of vastly different scales used to measure Taste and Nutrition? Read on.

Measurement Scales Matter!

The answer to the above Point to Ponder is Yes. It turns out that differences in the scales used to measure the data can have a significant impact on the results. It is not uncommon for different scales to be used in a research survey to measure different variables. In our fruit drink analysis, the measurement scale used to gauge customer taste reactions to our new beverage ranged from 1-50. The scale used to measure nutrition had a much narrower range from 1-5. The substantial difference in the ranges of these two measurement scales has a very definite impact on the cluster solution. In effect, the significantly wider range of the taste response scale results in taste being weighted more heavily than nutrition in the analysis. To see this, consider Figure 10.5 that plots the raw data using the same scale on the horizontal axis (i.e., for Taste Response) as on the vertical axis (i.e., for Nutrition Perception). In this figure, which reflects the impact of measurement scale, differences between clusters that are high vs. low on Nutrition Perception are obscured. From this figure it is easier to see why only 2 clusters were identified.

INSERT FIGURE 10.5 HERE

Another way to look at the problems that can be introduced by measurement scales is to consider the impact of scale on the calculated distance between two customers. For example, consider the distance between CUS3 and CUS7. The squared distance between these two customers, as noted in our dissimilarity matrix (see Table 10.3) is 1156.25. Notably, almost 99.9% ($[(46-12)^2] = 1156$) of this value is due to taste response, while hardly any is due to nutrition. The marketing manager must consider whether or not taste deserves to play such a dominant role in this squared distance computation. If there are not substantive reasons for the differences in the measurement scales to result in taste playing such a disproportionate role, one option is to rescale the data in order to equate the scales. In doing so, we remove scale differences as a principal explanation for our resulting cluster solution.

In our example, one way to equate the scales is to convert the taste response scale to a five point scale (identical to the scale used for nutrition) by dividing the collected taste values by 10. Using the rescaled values for taste response in order to compute the squared distance between CUS3 and CUS7, we now calculate the distance to be 11.81 ($[4.6-1.2]^2 + [4.5-4]^2$). A new summary of the taste response and nutrition data, using our rescaled taste response data, is provided in Table 10.7. If we create dissimilarity measures based on the data in Table 10.7 and then apply hierarchical clustering, a plot of the within-segment dissimilarity across the steps of the hierarchical clustering process is presented in Figure 10.6. The unusually large increase in heterogeneity going from the four cluster solution to the three cluster solution suggests that the optimal number of clusters is four as shown by the elbow (see Figure 10.6) – corresponding to what we initially determined based on our preliminary inspection of the data.

INSERT TABLE 10.7 HERE

INSERT FIGURE 10.6 HERE

> Point to Ponder: In the above example it was easy to make the scales similar as both the variables were perceptual in nature. What could you do to make scales similar when the variables are quite distinct, e.g., income and distance from work? Read on.

When variables are measuring quite distinct phenomena, they can be made comparable by standardizing them. To standardize a scale, we first subtract the sample mean from each score, and then divide the remainder by the sample's standard deviation. This process makes the mean 0 and the standard deviation 1 for all responses (Table 10.7 also provides standardized scores for taste and nutrition in the 5th and 6th columns). Regardless of approach, transforming or standardizing measurement scales can be a critical consideration in making the collected data suitable for analysis.

Statistical modeling is just one input into any decision regarding the appropriate number of clusters to retain. What should also guide the process is the ability to easily interpret the cluster solution, as well as the ability of the marketing manager to use the cluster solution to choose and implement an effective target market and marketing mix strategy. In some instances, the clustering methodology will not identify the same number of clusters as the manager's intuition. Care should always be taken to resolve any conflict between the manager's intuition and the empirical results.

In sum, hierarchical cluster analysis is an effective means of identifying market segments when the marketing manager makes no a priori assumption regarding exactly how many segments optimally describe the data. The dimensional distance between sample individuals, based on their responses to measures of attributes/characteristics important to the marketer, is used to form the clusters. The analysis starts from the premise that each individual customer is a cluster unto themselves, and follows a process of successive clustering iterations until only one cluster comprising all individuals remains. This hierarchical segmentation approach is useful when the number of subjects to be clustered is small, i.e., the sample size of respondents is small.

When the sample size is large, hierarchical clustering can become unwieldy. In this case, non-hierarchical clustering can be a viable alternative.

Point to Ponder: How would you calculate distances or similarities if the variables of interest are both discrete and continuous? Shouldn't you take some interesting electives to get such questions answered?

Non-Hierarchical Clustering

Non-hierarchical cluster analysis assumes that the number of clusters/segments is known and has been specified in advance. If the number of segments is unknown, multiple cluster solutions assuming different numbers of clusters can be developed. Management judgment can then be used to determine which cluster solution is most appropriate and actionable. Using our fruit beverage example, a process similar to that described below in Figure 10.7 would be followed. First, the marketing manager must specify the number of clusters. Let's assume that based on judgment and insight, four clusters are specified. Next, centroids, or seeds, for the four clusters must be selected. The seeds can be chosen based on a visual examination of the data, such as is provided in Figure 10.1, or the seeds can be chosen to represent some possible segments; or, they could be the values of four actual respondents who are judged to be typical of four different segments the manager might have in mind. After the four initial seeds have been designated, each of our twelve customers is assigned to one of the four clusters based on its distance from each of the four seeds. Each customer is assigned to the seed to which it is the closest.

Point to Ponder: What might a marketing manager consider when trying to come up with initial seeds? Might the selling organization have some ideas about the unique kinds of customers in the market? What about product design engineers? Could they be helpful here?

After all customers have been assigned to one of our four clusters, we next re-compute the four cluster centroids. Based on these new centroid values, we re-compute the distance from each customer to these centroids, and again assign (or reassign) them to the closest one. The cluster centroids (seeds) for our latest four cluster iteration are again re-computed. This iterative process continues until an optimum assignment of customers to clusters has been achieved. Typically, the iterations cease and an optimum cluster assignment is identified when the change in re-computed centroid values becomes either zero, or very small – less than some specified minimum amount.

INSERT FIGURE 10.7 HERE

Choice of Initial Seeds can be Critical

The effect of the initial selection of cluster seeds on the eventual cluster solution can be significant. A common rule of thumb is to select initial cluster seeds that are as far away from each other as possible. Once again, in our fruit beverage example, the data included in Table 10.7 from our hierarchical cluster analysis could have been helpful in choosing initial seeds. Given the power and affordability of computing these days, most of the popular statistical software packages that support non-hierarchical cluster analysis provide heuristics for starting seeds that insure that the final solution is quite robust.

Which Approach is Best?

Both hierarchical and non-hierarchical cluster analysis approaches have advantages and disadvantages. On the one hand, the hierarchical method requires no prior knowledge regarding how many clusters or segments best describe the data. On the other hand, hierarchical analysis using large data sets can require extensive computing resources, given the potential number of large dissimilarity matrices that must be computed and stored across clustering iterations. While the non-hierarchical approach may appear to be simpler, the number of clusters needs to be specified upfront. In many instances, the best approach may be to combine the use of hierarchical and non-hierarchical clustering. A smaller random sample of the full data set can be used to conduct the hierarchical analysis. The hierarchical solution provides the marketer with a better understanding of how many clusters and the initial seeds for the non-hierarchical solution with the complete data-set[ii].

> Point to Ponder: Both the methods discussed above assign respondents/cases/ customers to one and only one segment. Can you think of situations where a customer can belong to more than one segment? Can you also visualize "overlapping clusters"? Can you think of situations when this might occur? There are approaches that accommodate all these needs.

FACTOR ANALYSIS

In many market segmentation contexts, the substantial amount of useful information available to the marketer can be both a blessing and a curse. For example, consider the case of a toothpaste manufacturer attempting to introduce a new line extension to the market. From its qualitative, exploratory research (e.g., conducting one or more focus groups) the firm identifies a significant number of important product-related attributes that purport to influence consumers' perceptions and purchase intentions. These attributes are often expressed as benefits of using toothpaste products, such as the ability to prevent cavities, freshen breath, whiten teeth, remove tartar, prevent the buildup of plaque, promote healthy gums, protect sensitive teeth, taste good, etc. Based on these qualitative findings, quantitative research (often involving the collection of customer survey data) may then be used to gather numerical ratings of the importance of each of these different attributes according to a representative sample of potential customers. These customers may be asked in a survey to indicate their degree of agreement or disagreement, using a seven point scale (1=strongly disagree, 7=strongly agree), with statements such as "it is important to buy a toothpaste that freshens my breath," or "prevention of cavities is not an important benefit offered by a toothpaste." A data set of 20 hypothetical consumers' evaluations of toothpaste brands on 8 attributes which influence toothpaste purchase intention is presented in Table 10.8. The manager may wish to segment the market based on the stated importance of these eight different variables.

INSERT TABLE 10.8 HERE

Unfortunately, it is not uncommon for the marketer to face certain complications. One such complication is when certain attributes/variables in the analysis are highly inter-correlated (e.g., prevents cavities and removes tartar may be highly correlated). If there are four highly correlated variables that essentially represent one overall concept, and there is one other variable representing a different concept, then the four variables will have four times more influence on the overall data analysis, e.g., in cluster analysis, than is necessary or appropriate. Ideally, only one overall variable should represent the four highly correlated individual variables. Another complication is that in many instances the number of attributes available for use in developing market segments and segment profiles is simply unmanageable – it is so large that it must be reduced to a more reasonable number. Regarding both of these complications, one solution would be to limit the statistical analysis to a smaller selection of variables (attributes) – i.e., identifying and selecting only one of the four highly correlated variables , or intuitively choosing a smaller number of variables for the second case. Obviously, a major problem with this remedy is that it might be impossible to intuitively know which variables are correlated among themselves, and not with others. A better solution would be to reduce the number of variables by combining some of them into subsets or groupings of the total number of variables, and using these combined groupings of variables in the analysis. Factor analysis is a technique which can do this – collapsing a large set of variables into a set of smaller, necessary, and representative variables (factors).

Less is More

Factor analysis is a statistical method for identifying a smaller set of "factors" that capture the statistical information contained in a larger set of correlated variables. It is used to reduce and summarize data that otherwise would be either unwieldy due to the number of variables, or misleading due to the existence of significant collinearity between the measured attributes. Using a different example, consider the broad range of attributes on which a car manufacturer would want to collect detailed, relevant customer data. This list of attributes could number literally in the dozens, including engine size, horsepower, acceleration, fuel economy, seating capacity, storage room, towing ability, exterior styling, interior styling, upholstery options, cruise control, power doors/windows, stereo options, sticker price, and many more. Once again, the dilemma for the marketer is that this list presents too many attributes on which to base a reasonable set of segments. In addition, it is very likely that certain of these attributes are highly correlated, creating the possibility that a statistical analysis of their combined influence on a cluster analysis of the data is overly exaggerated. Factor analysis offers a potential solution to this dilemma. It is possible that a factor analysis of the many different variables embedded in the overall data could identify a smaller set of factors (e.g., performance, comfort, economy, luxury, convenience) on which to group the consumers and their preferences. In other words, factor analysis attempts to replace a large set of observed variables with a smaller set of new, unobserved variables; these new variables, or "factors," are used to develop interpretable, actionable segments and segment profiles.

The primary goal of factor analysis in our data-driven segmentation context is data reduction and summarization. The objective is to take the typically large number of variables of interest to the marketer, and then to examine and represent them as a reduced set of underlying factors. Conducting factor analysis starts with the marketer specifying the variables or attributes to be included in the analysis based on experience, insight and judgment. The mathematical process is based on an analysis of the correlations between these chosen attributes. For successful data reduction, there must be a certain degree of correlation between these attributes, which is most often the case. In our toothpaste example (see Table 10.9), we find relatively high correlations between certain attributes such as prevents cavities, removes tartar, prevents plaque buildup, promotes healthy gums, and protects sensitive teeth. Likewise, we find relatively high correlations between other attributes such as whitens teeth, freshens breath and tastes good. It would then seem that there are TWO underlying factors or variables. The fact that individuals want healthy teeth gives rise to identical answers to these particular questions (prevents cavities, removes tartar, prevents plaque buildup, promotes healthy gums, and protects sensitive teeth) about this construct. Similarly, the answers related to whitening, fresh breath, and tastes good might all be correlated because they stem from one underlying factor – cleanliness. So, besides the variables within a set being highly correlated to one another, we would expect that these same sets of variables would also be highly correlated with their respective underlying "factors." These correlations of individual variables with their underlying factors are called factor loadings. They are one of the key outputs of factor analysis and are used to interpret the meaning of the factors. Table 10.10 reports factor loadings for the toothpaste attribute data.

INSERT TABLE 10.9 ABOUT HERE

INSERT TABLE 10.10 ABOUT HERE

Point to Ponder: Correlations can be calculated with variables that are interval or ratio level. What would you do if some of the variables are discrete and factor analysis looked desirable? Take some electives ☺

Interpretation and Use

The interpretation / meaning attributed to some factor (unobserved) is a function of the specific variables (observed) that have high loadings on that factor. In other words, the rationale for interpreting a given factor as representing "promoting good dental health" is the fact that the five variables that have the highest loadings on this "factor" are: prevents cavities, removes tartar, prevents plaque buildup, promotes healthy gums and protects sensitive teeth. Unfortunately, it is not always a simple matter to interpret the factor analysis results. While it would be very convenient if each of the observed variables in a study only loaded on a single factor, it is not uncommon for certain variables (e.g., upholstery options in an automobile context) to load on two different and distinct factors (e.g., comfort and luxury). Moreover, some variables (product attributes) may not have particularly high loadings on any factor – because they are not highly correlated with any of the other variables used in the analysis. In such instances, variables may be added to the factor analysis solution in order to see if a factor captures this variable's correlation with other variables. Alternatively, these apparently unique variables may be deleted from the analysis. Deletion from factor analysis does not mean that this variable is not managerially important. It just means that it should be considered as a unique variable and can be used along with the identified factors, i.e., the new variables.

> Point to Ponder: How do you think variables could be combined? Think of a two dimensional graph where the x-axis represents one variable and the y-axis represents another. If the two variables are correlated, data on these variables will fall around a straight line. (Recall Scatter plot of a Regression Analysis.) Could that new straight line be thought of as a new axis that represents the new variable/factor?

Continuing with our example on toothpaste attributes, Table 10.10 shows the two factors with their loadings on the variables. The high loadings of variables – prevents cavities, removes tartar, prevents plaque buildup, promotes healthy gums and protects sensitive teeth, with factor 1 imply that factor 1 is really an attribute (factor) which concerns the medical and health related benefits of the toothpaste. Similarly, the high loadings of the variables – freshens breath, whitens teeth and tastes good, with factor 2 imply that factor 2 represents characteristics which are primarily cosmetic and focuses on the pleasures of using the toothpaste. Thus, this example shows that instead of working with eight variables, the marketer can actually work with only two. Factors can be thought of as “new variables” which, unlike their individual observed variables, cannot be directly observed in the data. However, in order to use a factor further in some subsequent marketing analysis, the marketer often needs to estimate a “score” for the factor. A *factor score* is just that – a value used to represent this newly combined set of variables (i.e., factor). While a simple average of the scores for each of the variables that load on that factor could be used to calculate a factor score, this option fails to utilize an important piece of information, namely, the individual factor loadings. Some individual variables are more highly correlated with a given factor than others and, hence, make more of a contribution to the new factor than the others. A better estimate of the factor score than a simple average would be to use a weighted average of the individual scores such that variables with higher factor loadings are weighted more heavily than variables with lower factor loadings. Although this weighted average provides a better estimate of the factor’s value, it’s important to remember that a factor score is never a perfect measure of the value for any unobserved factor.

The coefficients needed to calculate factor scores are different from the loadings. But suffice it to say that for our purposes, these factor loadings “give rise” to factor scores when combined with the values that a respondent reports for a toothpaste brand’s attributes. That is, respondent j’s score on factor 1 would be to some degree proportional to the following summation.

$\text{Factor1}_j \; \alpha \; .958\ V1_j - .079\ V2_j - .381\ V3_j + .886\ V4_j + .920\ V5_j + .871\ V6_j + .880\ V7_j - .189\ V8_j$,

where V1 – V8 are the values corresponding to the jth respondent and are available in Table 10.8.

Point to Ponder: How would you decide on the set of variables that would be factor analyzed? Just because some variables are correlated, do they then make good candidates for FA? The importance rating on cavity prevention might be correlated to political ideology but would you put this variable in the above factor analysis?

How Many Factors?

The number of factors to retain is again an issue here, as it was with the number of clusters to use in cluster analysis. (While cluster analysis grouped individuals based on their similarity of responses to questions, factor analysis groups variables.) Clearly there can be as many factors as the number of variables. But given the goal of reducing the number of variables, a lower number of factors than starting variables will always be the case with factor analysis. One way of deciding on the number of factors is to choose only those factors whose eigenvalue is greater than one. Eigenvalues measure the amount of variance that each factor explains. With our correlation matrix as input, each original variable is automatically scaled to have a variance of one, hence, choosing factors with eigenvalues greater than one makes sense – each new variable (factor) explains more variance than any of the individual, original variables.

Alternatively, similar to the mechanism used to determine the appropriate number of clusters, we can plot eigenvalues and look for what we called an elbow in cluster analysis but which we will call a "scree" for factor analysis. This technique plots the eigenvalues for each of the identified factors (see Figure 10.8). The slope of the change in eigenvalue from one factor to the next gives a feel for the amount of change in the variance explained by each factor. When that slope flattens out, as it does between factors 2 and 3, we infer that a two factor solution is optimal. Generally, the point at which the slope flattens out (the scree) denotes the optimum number of factors. As Figure 10.8 demonstrates, after two factors are extracted, only minimal amounts of explained variance are gained by extracting additional factors. Factors 1 and 2 explain 85.6% of the variance with Factor 1 explaining 69.7% and Factor 2 explaining 15.9%. It is pretty evident that in this Factor Analysis extraction of 2 factors is sufficient. Each of remaining factors 3 through 8 explain 5% or less of the variance.

INSERT FIGURE 10.8 HERE

Another workable approach is to identify the optimum number of factors based on some predetermined level of variance explained by the factor analysis solution. In other words, factors are extracted until the amount of cumulative variance explained reaches some acceptable level. In many marketing contexts, a benchmark of roughly 70 percent of the total variance explained is considered good. In addition, it is always possible that the number of factors can be specified *a priori* based on the knowledge and/or purposes of the marketer. In this case, the extraction of factors concludes when the predetermined number of factors has been reached.

> Point to Ponder: Why would one want to limit the number of factors to that number specified by the knowledge and/or purposes of the marketer? If an analyst ignored this advice and reported a solution with twice as many factors as suggested by the marketer, what do you think would happen?

Factor Rotation

As discussed above, an important output of the factor analysis is a set of factor loadings, which indicate the degree of correlation between the factors and the various variables. High correlations suggest that the variable and the factor are closely related, and vice versa. However, it is often the case that the initial factor analysis output is difficult to interpret because the factors, to varying degrees, are correlated with many different variables. In such cases, "rotation" of the factor matrix (the factor loadings for the various factors and variables) is helpful in simplifying and improving the interpretability of the results. In effect, rotation redistributes the variance explained by individual factors, such that each variable has significant loadings with only a few factors, preferably only one. A popular factor rotation method is the *varimax procedure*, which is designed to limit the number of variables that are highly correlated with a given factor – thus improving the marketer's ability to interpret the factor analysis solution.

> Point to Ponder: Why do you think that factor interpretability is important? Can cluster analysis be done with nard-to-interpret factors? If so, what problems are likely to arise in implementing the cluster solution?

FACTOR ANALYSIS AND CLUSTER ANALYSIS

Consider our earlier scenario where a manufacturer of personal computers is interested in segmenting the market. The firm collects information on eighteen different variables designed to reveal customers' attitudes toward personal computers (please refer back to Table 10.1). Designing a segmentation plan based on all of these variables would be a daunting task. Instead, the observed information can be first subjected to factor analysis in order to reduce the number of possible segmenting variables. From the eighteen variables presented in Table 10.1, factor analysis identifies three primary dimensions of customers' attitudes toward personal computers –

their knowledge of computers, their personal enthusiasm for computers, and their general negative regard for technology (as can be seen in Table 10.11). Factor scores can be calculated for each of these unobserved factors. Then, using these factor scores individuals can be grouped using hierarchical or non-hierarchical clustering techniques.

If the number of factors extracted in the factor analysis solution accounts for most of the variance in the data, we have accomplished the objective of data reduction without any significant cost – since the resulting factor scores should be very representative of the eighteen underlying variables. Also the smaller number of factors can result in a more stable cluster analysis solution. Finally, factor scores eliminate the problem where different scales with different ranges are used in the data collection. Using factor scores in the cluster analysis eliminates scaling concerns.

The results of the factor/cluster analysis for our computer example are presented in Table 10.12. Three factors (knowledge, personal enthusiasm, general negativity) have been extracted from the data and are used to represent the eighteen original variables included in our research survey. Using these three factors, the cluster analysis provides a seven cluster solution. (For the sake of easy interpretation the variable general negativity has be reversed to general positivity and signs appropriately flipped.) The seven discrete segments (not for me, selfish, converted, disgruntled, excitable, confused, opinionated,) are derived from the data. Members of each of these seven segments are similar to others within their own segments, and different from the members of the other six segments.

Segments 1 and 2 are polar opposites. Customers in Segment 1 ("Not-for-me") are not knowledgable and not enthusiastic about technology but see technology as being good for society, while customers in Segment 2 ("Selfish") know about technology, are personally enthusiastic butd see technology as being bad for society. Segment 3 ("Converted") shares Segment 2's knowledge and enthusiasm for technology and see technology as being good for society. Segment 4 ("Disgruntled Information Workers") is knowledgable but is not enthusiastic about technology and do not see technology as being good for society. It would seem that they work in the information processing area but really do not like it much. Segments 5 ("Excitable") and 6 ("Confused") are both enthusiastic about technology despite the fact that they aren't knowledgable but differ in their view of how good technology is for society. Finally, Segment 7 ("Opinionated") believes that technology is bad for society despite the fact that customers in this segment have little personal knowledge of or enthusiasm for technology.

> Pont to Ponder: Are there any of the identified segments to which you might have given a different name? What name would you have given? If you were a marketing manager, how would you choose among suggested names for segments? Does it matter what name we give?

The interpretability of this seven cluster solution would not be possible had we not initially reduced the eighteen original variables to three principal factors. Moreover, this segmentation solution, in all likelihood, could not have been determined creatively, using only the insight and experience of the marketer. It is only by examining and applying statistical techniques to the data that we are able to arrive at this interpretable and actionable segmentation plan.

INSERT TABLE 10.11 HERE

INSERT TABLE 10.12 HERE

The above examples illustrated how a marketing manager might go about segmenting a market explicitly based on one or more variables. In many statistical analyses, however, the homogeneity of respondents not only cannot be assumed, but degrees of heterogeneity in the sample may actually render some statistical approaches inappropriate and misleading. Thus, in certain modeling contexts, e.g., regression or conjoint analysis, estimating a single set of parameters or a single set of attribute importance ratings may fail to provide useful insights. This may be due to a common situation where the sample consists of more than one segment with each segment having its own set of parameters. For example, calculating the regression coefficient for the price sensitivity of a sample of respondents may result in a conclusion that price sensitivity is not a significant variable in understanding purchase intention. However, this conclusion could mask the fact that there are actually two distinct segments that comprise the sample – one that is very price sensitive and one that is not at all price sensitive. A statistical methodology called latent class regression simultaneously categorizes respondents into segments and estimates parameters for each segment.

LATENT CLASS ANALYSIS

To illustrate how latent class analysis works, consider the following hypothetical example. Data is collected from a sample of 50 consumers[iii]. The data collected includes smart phone usage and perceptions on five different variables. These variables are (V1) I work hard, (V2) I like to be on time, (V3) I always have my computer turned on, (V4) I think people work too hard, and (V5) I am interested in world affairs. In order to determine which of these variables has the greatest influence on smart phone usage, the marketer performs a regression analysis on the entire sample of responses and learns, as summarized in Table 10.13, that none of the 5 variables is significant in predicting customers' smart phone usage behavior.

INSERT TABLE 10.13 HERE

Supposing it was known that the data actually consisted of two segments – the first 25 respondents belong to the first segment and the second 25 respondents to a second segment. When separate regression analyses are performed for each segment of the overall sample (see Tables 10.14 and 10.15), we see that, for segment 1, (V1) I work hard, (V2) I like to be on time, (V3) I always have my computer turned on, and (V5) I am interested in world affairs are positively related to smart phone usage behavior and (V4) I think people work too hard is negatively related. All 5 variables are significant except V2. Whereas, in Segment 2 the relationship is exactly opposite (i.e., the signs of the coefficients are opposite to those in Segment 1) and also all 5 variables are significant. In sum, from a regression of the aggregate sample which showed no variables were significant, we arrive at a situation that most of the variables are significant in determining smart phone usage but in opposite ways for the 2 different segments. Clearly, the aggregate analysis is misleading.

The above illustration assumed that we knew that there were two segments, and we also knew who belonged to which segment. Generally, however, the marketer has no *a priori* sense of what and how many segments actually describe the market, and which consumers belong to which segments. In that case, latent class analysis is used. Tables 10.16 and 10.17 show the result of using this statistical procedure on the entire sample of 50 respondents. Two, 3 and 4-

segment models are estimated. Considering the Log Likelihood associated with each solution (higher – less negative – Log Likelihood indicating a better fit) reported in Table 10.16, the two segment model is chosen. Without getting bogged down in the statistical details, we can see that the criteria stop changing much going beyond two segments. Just like the elbow in clustering (Figure (10.6) and the scree plot in factor analysis (Figure 10.8), here too we use the "elbow" criterion and this criterion suggests a two segment solution.

INSERT TABLE 10.14 HERE

INSERT TABLE 10.15 HERE

INSERT TABLE 10.16 HERE

Table 10.17 provides the coefficients for the two segment latent class model (K=2). Comparing the coefficients from latent class analysis in Table 10.17 to those obtained in separate analyses in Tables 10.14 and 10.15, we see that latent class regression analysis reproduces the two underlying segments and their response coefficients very well.

INSERT TABLE 10.17 HERE

OTHER SEGMENTATION PROCEDURES: SEGMENTATION TREES

While there are literally hundreds of different clustering procedures and algorithms, each having been developed for a very specific purpose, a class of procedures called segmentation trees is quite popular. Consider the example of a long distance telephone service supplier who is interested in segmenting users based on their monthly long distance calling expenditures.[iv] Data has been collected that allows the marketer to examine how monthly long distance spending (the dependent variable) varies across a set of predictor variables such as income, education, home ownership, etc. Automatic Interaction Detection (AID) is a method that uses this type of data to systematically and successively split the total sample into mutually exclusive segments; each split of the sample is based on identifying the particular predictor variable (e.g., family income), and level of the variable (e.g., low income versus medium or high income) that *best* explains the variation in the dependent variable (e.g., monthly long distance expenditures). The output of the process is a segmentation tree that visually and numerically describes the makeup of segments that constitute an overall sample of data.

INSERT FIGURE 10.9 HERE

The figure shows the result of applying the Automatic Interaction Detection (AID) technique to develop segmentation trees. The average total spending by all respondents is $20.05. From all the variables (income, home ownership, education, local or long distances moves) and their levels, the first segmenting variable is income (see Figure 10.9). Families with annual income less than $40,000 spend $6.92 per month on long distance, while families earning $40,000 or more spend $41.35 on average. This split is the top-most because with this split the within group homogeneity and between group heterogeneity is the maximum as compared to all other possible splits. The lower income segment represents 62% of the market, while 38% of the market is in the higher income segment. Now, each segment (high income, low income) is treated as if it was a new sample, and the process is repeated. For lower income families, the variable that ***next*** best explains variation in monthly long distance expenditures is the education level of the head of household. As binary split of the lower income families shows that households whose heads have less than a college education spend $6.21 per month on long distance, whereas households whose heads have college education or more spend nearly twice that amount ($11.05). These two segments (lower income with < college education and lower income with > college education)) represent 52% and 10% respectively of the total sample. Note that for families with income greater than $40,000, a different predictor variable (home ownership) is used to split the data. This is because home ownership (owning versus renting) apparently optimizes the criterion (that is based on within and between group variances) in monthly long distance expenditures for higher income families more than any other predictor variable (such as education level of head of household). As Figure 10.9 illustrates, higher income families that own their homes spend less ($39.67) on long distance than higher income families who rent ($52.39). The AID process continues by selecting those remaining predictor variables that explain the largest amount of variation in the dependent variable – for each previously defined split. In this example, family income, education level of head of household, home ownership, and location of previous move are the most important predictors of variation in monthly long distance expenditures. Five different segments are identified through using this AID algorithm. The segmentation tree that results from the process can be very helpful in providing insights regarding segments and corresponding marketing mix implications. The segment with the highest level of monthly long distance spending ($52.39) includes households that earn more than $40,000 and rent their homes. Note, however, that this segment of heavy long distance spenders represents only five percent of the total sample.

Point to Ponder: Why do you think that higher income households who are renting their homes spend the most on long distance? Why do you think that low income families in which the head of household attended college spend more on long distance than do low income families whose head of household did not attend college? What marketing interventions can you come up with to exploit the insight provided by this segmentation scheme?

The binary AID procedure is limited to splitting data dichotomously. Each split is simply a binary split of the data. As described above, the AID process examines all possible *two-way* splits of each segment, for each variable, and uses the split that explains the most variation in the dependent variable (monthly long distance expenditures). However, AID cannot handle situations involving categorical dependent variables. Alternative segmentation tree methods, such as CHAID (Chi-square Automatic Interaction Detection), may be appropriate when using these types of dependent variables. The CHAID approach is applicable for all types of variables since variables measured using interval scales can always be converted into categorical variables by dividing the range of responses into sensible categories. In addition to the flexibility of CHAID in handling different types of dependent variables, another benefit is that the splits of the data do not have to be binary. As such, CHAID is a more commonly used segmentation tree technique. Regardless of which approach is used, segmentation tree procedures can offer the marketer terrific insights regarding the particular characteristics that best describe potential market segments, and that best distinguish them from other potential segments.

PROFILING SEGMENTS

Once the segments are defined according to one or more segmentation bases (*a priori* or empirical), and respondents are assigned to the resulting groups, the marketer must focus on profiling each segment in order to better understand its distinctive characteristics. The profiling variables are different than the variables that are the bases of segmentation. To return to our computer example, the bases of segmentation were knowledge and attitude type variables. Profiling the seven segments could imply relating each segment to the following kinds of variables:

- Do different market segments vary in terms of the types and amounts of media they consume?
- Are any of the segments more brand loyal than other? And if so, to which brands?
- Are any segments more price-sensitive than others?

Profiling is always very useful when segments are created *a priori*. Thus, one might segment the market into brand loyal and not loyal individuals; price sensitive and not so; light users and heavy users; etc. In each of these cases, we would like to know what are the other characteristics of each segment, e.g., how do brand loyal individuals differ from their counterparts in terms of demographic, psychographic, lifestyle, media consumption habits, etc.

Profiling segments is invaluable in developing effective product, price, distribution, and promotion strategies. Discriminant Analysis is one methodology that can help the marketer to profile segments.

PROFILING SEGMENTS USING DISCRIMINANT ANALYSIS

Once the marketer has identified different groups of respondents based on their responses to various measures, or based on *a priori* criteria, important questions still remain. Are the various segments significantly different from each other on other variables of interest? On which particular variables are the groups most different? Can I use these variables to predict segment membership for any other given individual? Discriminant analysis is a method for attempting to answer these questions. It can be considered to be a statistical technique for analyzing data when the dependent variable is categorical, a la regression analysis. In fact, discriminant analysis is identical to regression for dichotomous independent variables, but it is different when examining categorical dependent variables with more than two levels (e.g., three different age ranges – 18-35 years old, 36-55 years old, over 55 years old). Unlike cluster analysis, where the membership of individuals in groups is not previously known, with discriminant analysis membership in groups is known. The basic output of the analysis is a set of discriminant functions – linear combinations of the independent/predictor variables that best discriminate between the categories (two or more) of the dependent variable. There will always be one less (K-1) discriminant function than there are levels (K) of the categorical dependent variable. However, not all discriminant functions estimated may be statistically significant. Interpretation of the discriminant functions is similar to that of factor analysis. The segments differ the most on the dimension reflected by the first Discriminant function. The second Discriminant function is the second most important dimension on which the segments differ; and so forth.

To illustrate the value of Discriminant analysis, take the example of a firm promoting a health food product. The firm collects data on a number of variables purporting to indicate a person's degree of health consciousness (see Table 10.18). Four types of individuals are known to exist: those who are (1) very conscious, (2) somewhat conscious, (3) somewhat not conscious, and (4) not conscious about their health. These four categories represent the four different levels of the categorical dependent variable.

Because four levels exist, we can estimate three (K-1) discriminant functions.

Discriminant Functions

$$Y_1 = .9X_1 + .8X_2 + .75X_3 + .2X_4 + .1X_5 + .05X_6 + .1X_7 + .2X_8 + .01X_9$$
$$Y_1 = .09X_1 + .01X_2 + .1X_3 + .7X_4 + .7X_5 + .85X_6 + .1X_7 + .2X_8 + .01X_9$$
$$Y_1 = .3X_1 + .4X_2 + .3X_3 + .2X_4 + .1X_5 + .5X_6 + .5X_7 + .3X_8 + .1X_9$$

•*First two significant; third is not.*

It turns out that only two of the three estimated functions are significant. The first Discriminant function can be interpreted as the eating dimension because the coefficients that are large load on to variable X_1, X_2, and X_3 the three measured variables on diet. Or in other words, the most difference between the 4 groups is in the way they take care of their diet. The second function can be interpreted as the exercise dimension, i.e., the 4 groups next most differ in terms of how they exercise. This interpretation stems from the high coefficients of variables X_4, X_5, and X_6. Similar to regression analysis, the value of the various discriminant coefficients depends on the other predictors (variables) included in the discriminant analysis.

> Point to Ponder: Factor Analysis of the kind described earlier in the chapter does not allow for discrete variables. What do you think is the case with Discriminant Analysis? Would categorical predictor variables work here? Hint: Think regression with categorical predictor variables.

Discriminant analysis can then also be used for classification purposes. A new respondent can be assigned to a group using the classification functions that are an output of the

Discriminant analysis process. It is beyond the scope of this book to get into the details. Discriminant Analysis references listed at the end of this chapter are excellent sources for learning more.

> Point to Ponder: Factor Analysis requires variables to be correlated for any meaningful use of the technique. Regression on the other hand can be severely limited by the predictor variables being correlated (ala multicollinearity). What do you think is the role of correlated predictor variables in the Discriminant Analysis setting?

FUTURE DIRECTIONS

In this chapter we have discussed data driven segmentation methods. Market segmentation which leads division of market into distinct groups of potential buyers who have similar needs and wants helps in identifying bases for segmenting the market and develop profiles of segments. This leads to selection of segments to enter, develop measures of attractiveness and select target segment(s). Finally, it helps in formulation of competitive positioning and marketing mix, develops positioning for each target segment and develops marketing mix for each target segment.

It is almost impossible to discuss the future directions in which data-driven segmentation will evolve. The reason for this is that this area is totally dependent on statistical techniques. Newer techniques are being constantly developed and there are academic journals and books devoted to such statistical techniques.

KEY TAKEAWAYS

- Cluster Analysis mainly deals with minimization of distance functions; it can either be hierarchical or non-hierarchical. The drawback of this approach is that it is non-probabilistic, heuristic (not model based and does not provide an optimal solution) and tandem approaches are needed.
- The similarity measures used in cluster analysis are scale-dependent. It is often a good idea to standardize the data representing key marketing variables before conducting the cluster analysis.
- Remember, with non-hierarchical clustering, the process may be sensitive to the choice of initial cluster seeds. Different starting seed values can lead to different cluster solutions. It might be worth trying different values of seeds.
- Be careful of the possible correlation between different variables used in the analysis. If two of the variables are not independent of each other (e.g., house-size, income), the analysis will tend to grant them a higher weighting in forming the clusters – in effect it's as if they are being counted twice. This may result in an analysis that disguises the importance of other variables in explaining the market's structure and its resulting target market and marketing mix implications.
- Evaluating the *reliability* of any cluster analysis solution is a necessary step in the segmentation process. Reliability refers to the degree to which similar cluster solutions will be obtained across different data samples from the same population. Using a randomly selected holdout sample to examine cluster solution reliability is a common approach. The holdout sample is assigned to clusters using the cluster solution arrived at by the estimation sample. Then, a cluster analysis is performed for the entire sample. The degree of agreement between the

customer assignments for the holdout sample and the entire sample is generally an indicator of cluster solution reliability.

- The *validity*, or interpretability, of the cluster solution is also key. The cluster solution must make sense to the marketing manager. In many cases, the *external validity* of the cluster solution can only be evaluated after the segmentation plan is implemented and purchase data have been collected. However, the *face validity* of the solution must be evaluated in order to assure that the clusters identified are meaningful, are related to real, actionable marketing factors and variables, and are able to be operationalized by the marketing manager in terms of a target market and marketing mix strategy.
- The cluster analysis methods discussed here assume that a given customer can only belong to a single cluster. In certain marketing applications, this condition is not practical or appropriate. So, some methods allow for customers to belong to more than one segment and they estimate the "proportion" of the customer that belongs to a segment. Some methods, though assuming that a customer can belong to only one segment, cannot deterministically categories the customer to that segment. In that case, the customer is assigned to a number of segments with a probability associated with each assignment (Grover and Srinivasan 1987)[v].
- Moreover, the clustering methods discussed above require either ratio-scaled or interval-scaled data. Certainly marketing situations exist where other types of data such as ordinal-scaled (e.g., income in categories) or categorical (e.g., gender) variables would be appropriately used. Newer techniques, while not the focus of this discussion, are available to the marketer is such instances (Kamakura and Russell, 1989)[v].
- Once again, cluster analysis is no substitute for managerial judgment and experience. It should be viewed as an important *input* to the marketing manager's decision making process, not

a replacement for it. Cluster analysis doesn't by itself provide answers, but instead provides a means of examining and analyzing data that better prepares the manager to make effective segmentation decisions. Many different types of characteristics or variables, such as age, income, marital status, personality types, attitudes, attribute importance ratings, and lifestyle can be used to segment markets. The appropriate variables/characteristics are not determined by the cluster analysis methodology, but rather are determined by the judgment, insight and experience of the marketing manager.

- Factor Analysis: This is mainly a method of data reduction where we identify underlying factors which are correlated to the variables in question. This results in computational simplicity in identifying segments. Combining cluster analysis and factor analysis definitely gives more intuitive market segmentation results.
- The payoff from using the factor-cluster analysis approach is dependent on whether the specific factors used in the cluster analysis are helpful in differentiating between segments. Just like any variable, a factor too might not be significant in differentiating between segments. This can be sensed if the factor has almost the same mean value in all the segments.
- Latent Class Mixture models: This is the most efficient way of market segmentation since it uses powerful probabilistic techniques and simultaneously identifies segments and estimates model parameters in each of the segments. However, one of the significant drawbacks of this method is the possibility of local optima and assumes existence of segments.

Table 10.1

Attitudes toward Computers and Technology

My friends think of me as a knowledgeable source of information about software
I am enthusiastic about Consumer Electronics/ Digital Devices
I am excited about Computers
I seek out new ways to use my PC to its full potential
I seek out new ways to integrate the PC with other devices
I am excited about the internet
New technology has a positive impact on my life
I love to try new things
I continuously engage in learning
I am already taking advantage of new technologies that enable me to share experiences with my friends and family
I often turn to my PC or other technology or application to solve a home or business problem
My friends think of me as a knowledgeable source of information about the internet
I go out of my way to stay on top of the latest developments in technology
I use a computer because I feel I have to, not because I enjoy it
I feel that computers and technology are changing much too quickly
I couldn't imagine life without my PC
I usually try new software before my friends and coworkers do
I think computers are bad for society

Table 10.2
Fruit Drink Customer Profile

Customer	Taste Response	Nutrition Perception
CUS1	42	5
CUS2	44	4
CUS3	46	4.5
CUS4	45	1
CUS5	39	1
CUS6	43	2
CUS7	12	4
CUS8	16	4
CUS9	14	3
CUS10	18	1.5
CUS11	22	1
CUS12	26	1.5

Table 10.3
Dissimilarity Matrix for Fruit Drink Data

Consumer	Consumer											
	CUS1	**CUS2**	**CUS3**	**CUS4**	**CUS5**	**CUS6**	**CUS7**	**CUS8**	**CUS9**	**CUS10**	**CUS11**	**CUS12**
CUS1	0	5	16.25	25	25	10	901	677	788	588.25	416	268.25
CUS2	5	0	4.25	10	34	5	1024	784	901	682.25	493	330.25
CUS3	16.25	4.25	0	13.25	61.25	15.25	1156.25	900.25	1026.25	793	588.25	409
CUS4	25	10	13.25	0	36	5	1098	850	965	729.25	529	361.25
CUS5	25	34	61.25	36	0	17	738	538	629	441.25	289	169.25
CUS6	10	5	15.25	5	17	0	965	733	842	625.25	442	289.25
CUS7	901	1024	1156.25	1098	738	965	0	16	5	42.25	109	202.25
CUS8	677	784	900.25	850	538	733	16	0	5	10.25	45	106.25
CUS9	788	901	1026.25	965	629	842	5	5	0	18.25	68	146.25
CUS10	588.25	682.25	793	729.25	441.25	625.25	42.25	10.25	18.25	0	16.25	64
CUS11	416	493	588.25	529	289	442	109	45	68	16.25	0	16.25
CUS12	268.25	330.25	409	361.25	169.25	289.25	292.25	106.25	146.25	64	16.25	0

Table 10.4
Fruit Drink Example
Eleven-Cluster Solution

Cluster Number	Cluster Label	Cluster Membership	Taste Reaction	Nutrition Perception
1	CUS1	CUS1	42	5
2	CUS4	CUS4	45	1
3	CUS5	CUS5	39	1
4	CUS6	CUS6	43	2
5	CUS7	CUS7	12	4
6	CUS8	CUS8	16	4
7	CUS9	CUS9	14	3
8	CUS10	CUS10	18	1.5
9	CUS11	CUS11	22	1
10	CUS12	CUS12	26	1.5
11	CLUS11	CUS2,CUS3	45	4.25

Table 10.5
Fruit Drink Example
Dissimilarity Matrix for Eleven Cluster Solution

Cluster	Cluster										
	CUS1	**CUS4**	**CUS5**	**CUS6**	**CUS7**	**CUS8**	**CUS9**	**CUS10**	**CUS11**	**CUS12**	**CLUS11**
CUS1	0	25	25	10	901	677	788	588.25	416	268.25	9.56
CUS4	25	0	36	5	1098	850	965	729.25	529	361.25	10.56
CUS5	25	36	0	17	738	538	629	441.25	289	169.25	46.56
CUS6	10	5	17	0	965	733	842	625.25	442	289.25	9.06
CUS7	901	1098	738	965	0	16	5	42.25	109	202.25	1089.06
CUS8	677	850	538	733	16	0	5	10.25	45	106.25	841.06
CUS9	788	965	629	842	5	5	0	18.25	68	146.25	962.56
CUS10	588.25	729.5	441.25	625.25	42.25	10.25	18.25	0	16.25	64	736.56
CUS11	416	529	289	442	109	45	68	16.25	0	16.25	539.56
CUS12	268.25	361.25	169.25	289.25	202.25	106.25	146.25	64	16.25	0	368.56
CLUS11	9.56	10.56	46.56	9.06	1089.06	841.06	962.56	736.56	539.56	368.56	0

Table 10.6
Fruit Drink Example
Clustering Summary for Hierarchical Clustering

Step	Number of Clusters	Clusters Merged		Name of new cluster	Size of new cluster	Subjects in the new cluster	Dissimilarity
1	11	CUS2	CUS3	CLUS11	2	CUS2, CUS3	4.250
2	10	CUS8	CUS9	CLUS10	2	CUS8, CUS9	5.000
3	9	CUS4	CUS6	CLUS9	2	CUS4, CUS6	5.000
4	8	CLUS11	CLUS9	CLUS8	4	CUS2, CUS3, CUS4, CUS6	8.563
5	7	CUS7	CLUS10	CLUS7	3	CUS7, CUS8, CUS9	9.250
6	6	CUS1	CLUS8	CLUS6	5	CUS1, CUS2, CUS3, CUS4, CUS6	10.766
7	5	CUS11	CUS12	CLUS5	2	CUS11, CUS12	16.250
8	4	CLUS7	CUS10	CLUS4	4	CUS7, CUS8, CUS9, CUS10	20.694
9	3	CLUS6	CUS5	CLUS3	6	CUS1, CUS2, CUS3, CUS4,CUS5, CUS6	30.290
10	2	CLUS4	CLUS5	CLUS2	6	CUS7, CUS8, CUS9, CUS10, CUS11, CUS12	84.516
11	1	CLUS3	CLUS2	CLUS1	12	CUS1, CUS2, CUS3, CUS4,CUS5, CUS6, CUS7, CUS8, CUS9, CUS10, CUS11, CUS12	633.535

Table 10.7
Fruit Drink Customer Profile
Rescaled Data

Customer	Taste Response	Nutrition Perception	Taste Response Rescaled	Taste Response Standardized	Nutrition Perception Standardized
CUS 1	42	5	4.2	0.83	1.50
CUS 2	44	4	4.4	0.98	0.84
CUS 3	46	4.5	4.6	1.12	1.17
CUS 4	45	1	4.5	1.05	-1.12
CUS 5	39	1	3.9	0.61	-1.12
CUS 6	43	2	4.3	0.91	-0.46
CUS 7	12	4	1.2	-1.36	0.84
CUS 8	16	4	1.6	-1.06	0.84
CUS 9	14	3	1.4	-1.21	0.19
CUS 10	18	1.5	1.8	-0.92	-0.79
CUS 11	22	1	2.2	-0.63	-1.12
CUS12	26	1.5	2.6	-0.33	-0.79

Table 10.8

Attributes of Toothpaste:

Attribute	Variable
Prevents cavities	V1
Removes tartar	V2
Prevents plaque buildup	V3
Promotes healthy gums	V4
Protects sensitive teeth	V5
Whitens teeth	V6
Freshens breath	V7
Tastes good	V8

Table 10.9
Correlation Matrix of Toothpaste Attributes

Variables	***V1***	***V1***	***V3***	***V4***	***V5***	***V6***	***V7***	***V8***
V1	1.00							
V1	**0.80**	1.00						
V3	**0.90**	**0.81**	1.00					
V4	**0.79**	**0.69**	**0.86**	1.00				
V5	**0.81**	**0.81**	**0.73**	**0.73**	1.00			
V6	-0.03	-0.06	0.09	0.11	0.02	1.00		
V7	-0.42	-0.19	-0.18	-0.15	-0.21	**0.58**	1.00	
V8	-0.07	-0.06	0.05	-0.08	-0.05	**0.72**	**0.62**	1.00

Table 10.10

Factor Loadings (2 – Factor Solution)

	Factor	
	1: Promotes good dental health	**2: Freshens mouth**
Prevents cavities	.958	-.024
Removes tartar	.886	.078
Prevents plaque buildup	.920	.192
Promotes healthy gums	.871	.170
Protects sensitive teeth	.880	.128
Whitens teeth	-.381	.772
Freshens breath	-.079	.884
Tastes good	-.189	.872

Table 10.11

	Factors		
Determinants of Personal Computer Purchase Intentions	**1 Technology Knowledge**	**2 Personal Enthusiasm for Technology**	**3 General Negativity**
My friends think of me as a knowledgeable source of information about software	**.823**	**.266**	**-.130**
I am enthusiastic about Consumer Electronics/ Digital Devices	**.717**	**.431**	**-.149**
I am excited about Computers	**.576**	**.593**	**-.199**
I seek out new ways to use my PC to its full potential	**.593**	**.608**	**-.087**
I seek out new ways to integrate the PC with other devices	**.735**	**.403**	**-.047**
I am excited about the internet	**.279**	**.766**	**-.196**
New technology has a positive impact on my life	**.381**	**.783**	**-.185**
I love to try new things	**.323**	**.797**	**-.092**
I continuously engage in learning	**.270**	**.782**	**-.006**
I am already taking advantage of new technologies that enable me to share experiences with my friends and family	**.553**	**.663**	**-.059**
I often turn to my PC or other technology or application to solve a home or business problem	**.594**	**.567**	**-.075**
My friends think of me as a knowledgeable source of information about the internet	**.764**	**.427**	**-.151**
I go out of my way to stay on top of the latest developments in technology	**.775**	**.304**	**-.015**
I use a computer because I feel I have to, not because I enjoy it	**-.066**	**-.133**	**.715**
I feel that computers and technology are changing much too quickly	**-.198**	**.086**	**.740**
I couldn't imagine life without my PC	**.398**	**.321**	**-.195**
I usually try new software before my friends and coworkers do	**.834**	**.169**	**-.043**
I think computers are bad for society	**.028**	**-.226**	**.707**

Table 10.12
Factor Plus Cluster Analysis Solution

	1 Not-for-me	2 Selfish	3 Converted	4 Disgruntled Information Worker	5 Excitable	6 Confused	7 Opinionated
Size	17%	12%	20%	8%	15%	10%	17%
Knowledge	-.35	+.90	+1.06	+.83	-.86	-.88	-.63
Personal Enthusiasm	-1.12	+.60	+.46	-1.11	+.62	+1.21	-.60
General Positivity	.84	-1.12	+.68	-.60	+.80	-.59	-.93

Table 10.13

Regression Model Using All 50 Consumers

Dependent Variable: Smart Phone Usage
R^2=.06

Predictor Variable	Estimated Coefficient	t
I work hard	-1.6	-.87
I like to be on time	.35	.21
Always have my computer on	1.93	1.1
People work too hard	.49	.23
I am interested in world affairs	1.64	.93
Constant	.86	.49

Table 10.14

Regression Model Using First 25 Consumers

Dependent Variable: Smart Phone Usage
R^2=.99

Predictor Variable	Estimated Coefficient	t
I work hard	1.5	5.4
I like to be on time	.37	1.7
Always have my computer on	7.5	34.8
People work too hard	-5.8	-18.8
I am interested in world affairs	1.9	8.1
Constant	.31	1.3

Table 10.15

Regression Model Using Second 25 Consumers

Dependent Variable: Smart Phone Usage
R^2=.99

Predictor Variable	**Estimated Coefficient**	**t**
I work hard	-1.3	-6.0
I like to be on time	-.51	-2.3
Always have my computer on	-7.5	-35.8
People work too hard	6.3	25.5
I am interested in world affairs	1.8	-8.0
Constant	-.15	-.67

Table 10.16

Log Likelihoods for 1 to 4 Segments

Number of Segments Estimated	**Log Likelihood**
1	-191
2	-99
3	-92
4	-85

Table 10.17

Latent Class Regression Using All 50 Consumers

Dependent Variable: Smart Phone Usage
R^2=.99

Predictor Variable	Estimated Coefficient Segment 1	Estimated Coefficient Segment 2
I work hard	1.6	-1.2
I like to be on time	.36	-.53
Always have my computer on	7.5	-7.6
People work too hard	-5.9	6.3
I am interested in world affairs	1.9	-1.8
Constant	.29	-.18

Table 10.18

Discriminant Analysis
Hypothetical Variables

Variable Name	Variable Description
X1	I am careful about the amount I eat
X2	I am careful about when I eat
X3	I am careful about what I eat
X4	I try getting exercise every day
X5	I have a physically intensive job
X6	I am an outdoor activities kind of person
X7	I routinely go for an annual check-up
X8	I take vitamins every day
X9	I read health magazines

Figure 10.1

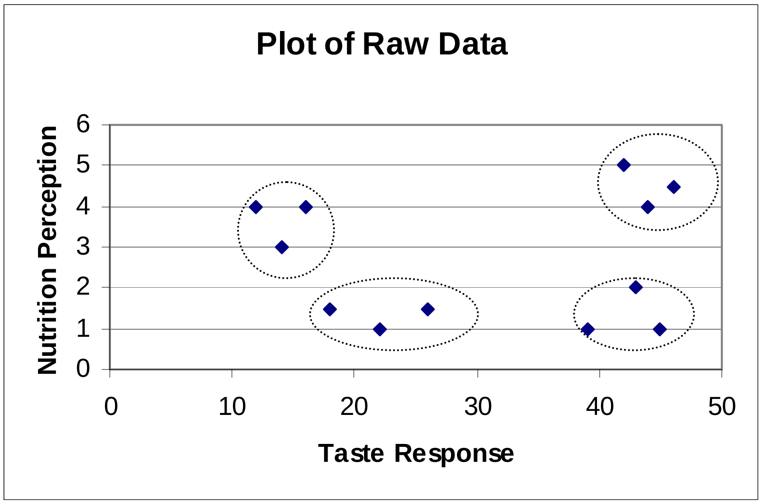

Plot of Raw Data
Nutrition Perception
6
5
4
3
2
1
0
0
10
20
30
40
50
Taste Response

Figure 10.2
Distance between Two Clusters for Centroid, Single Linkage and Complete Linkage Methods

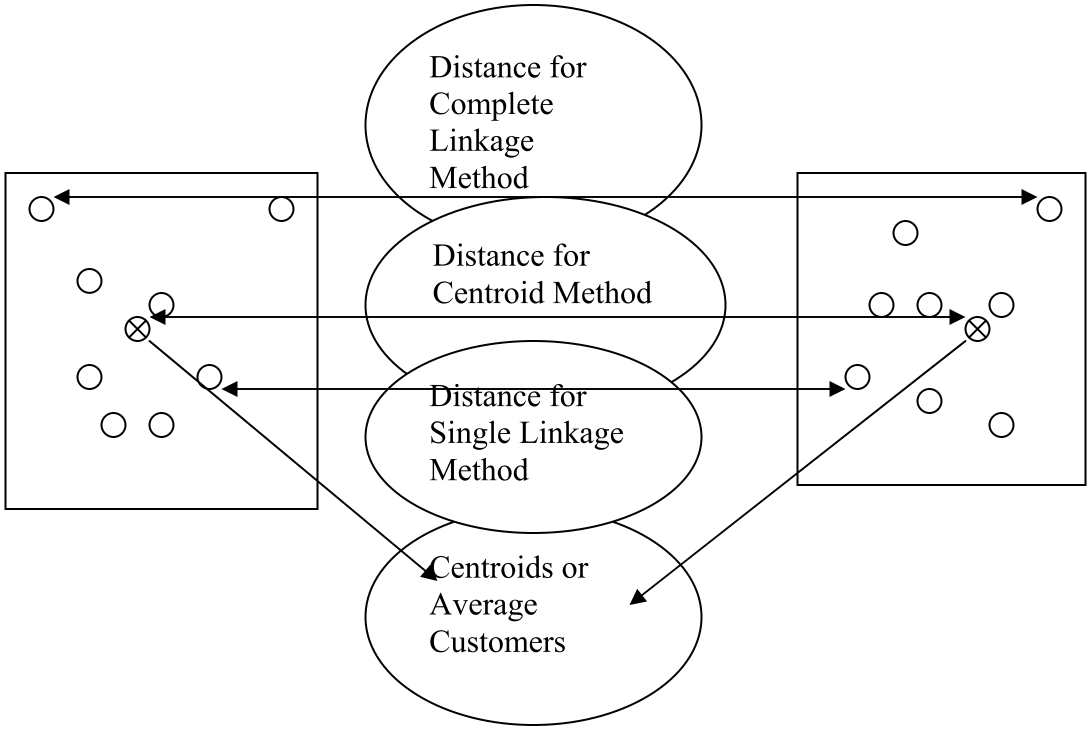


Figure 10.3 Dendogram Illustrating Hierarchical Clustering Solution

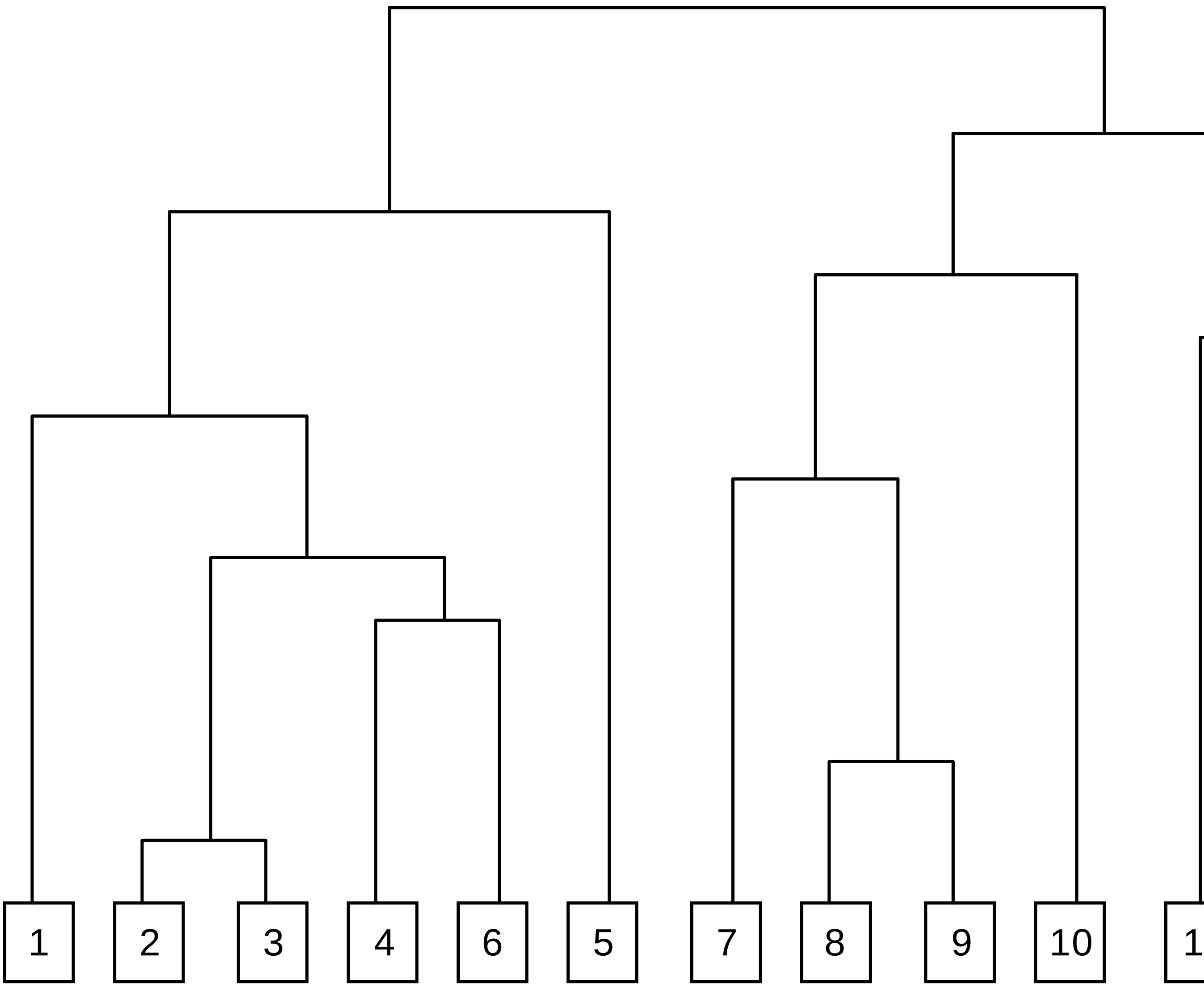


Figure 10.4

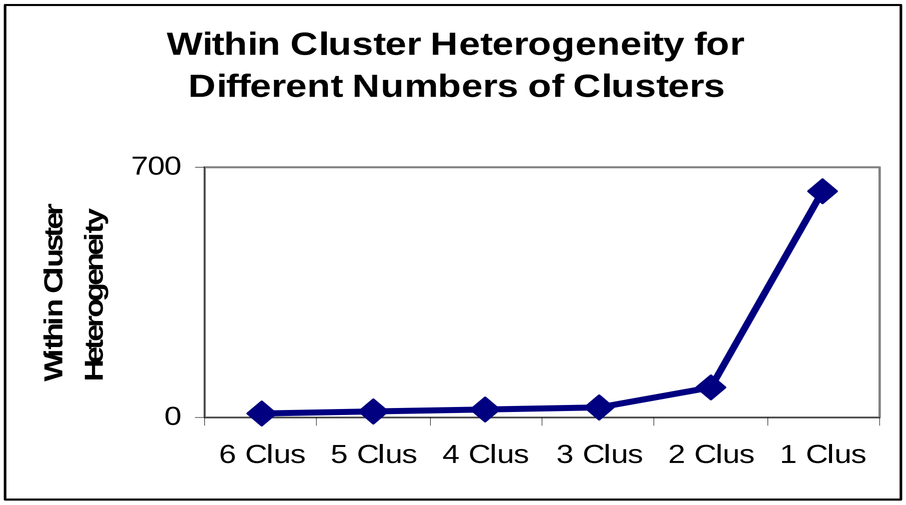
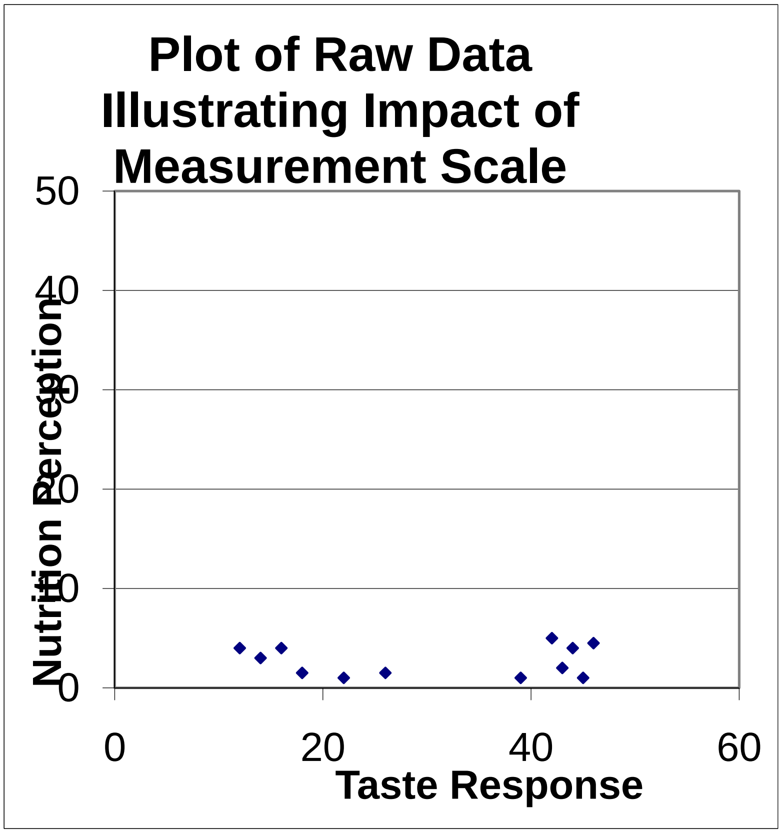

Within Cluster Heterogeneity for Different Numbers of Clusters
Within Cluster Heterogeneity
700
0
6 Clus
5 Clus
4 Clus
3 Clus
2 Clus
1 Clus

Figure 10.5



Plot of Raw Data Illustrating Impact of Measurement Scale
50
40
30
20
10
0
0
20
40
60
Nutrition Perception
Taste Response

FIGURE 10.6

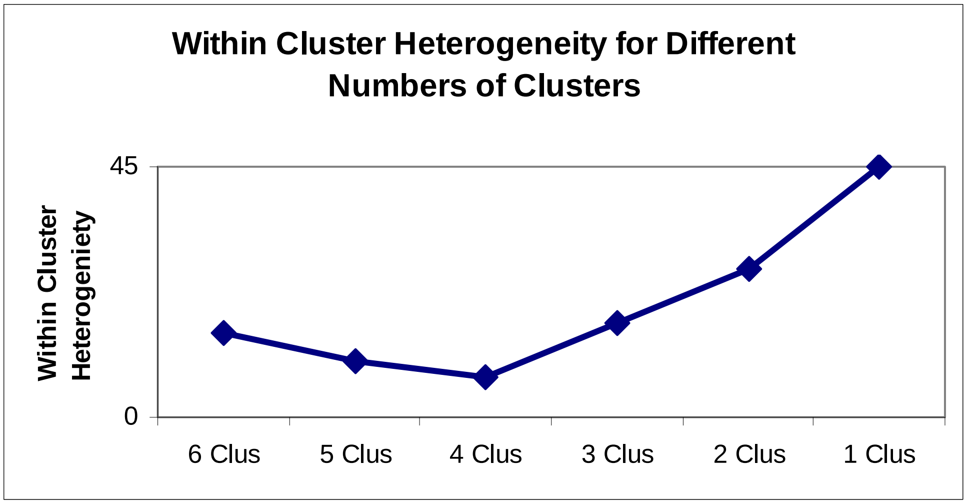

Within Cluster Heterogeneity for Different Numbers of Clusters
45
0
Within Cluster Heterogeneity
6 Clus
5 Clus
4 Clus
3 Clus
2 Clus
1 Clus

Figure 10.7
Non-Hierarchical Clustering

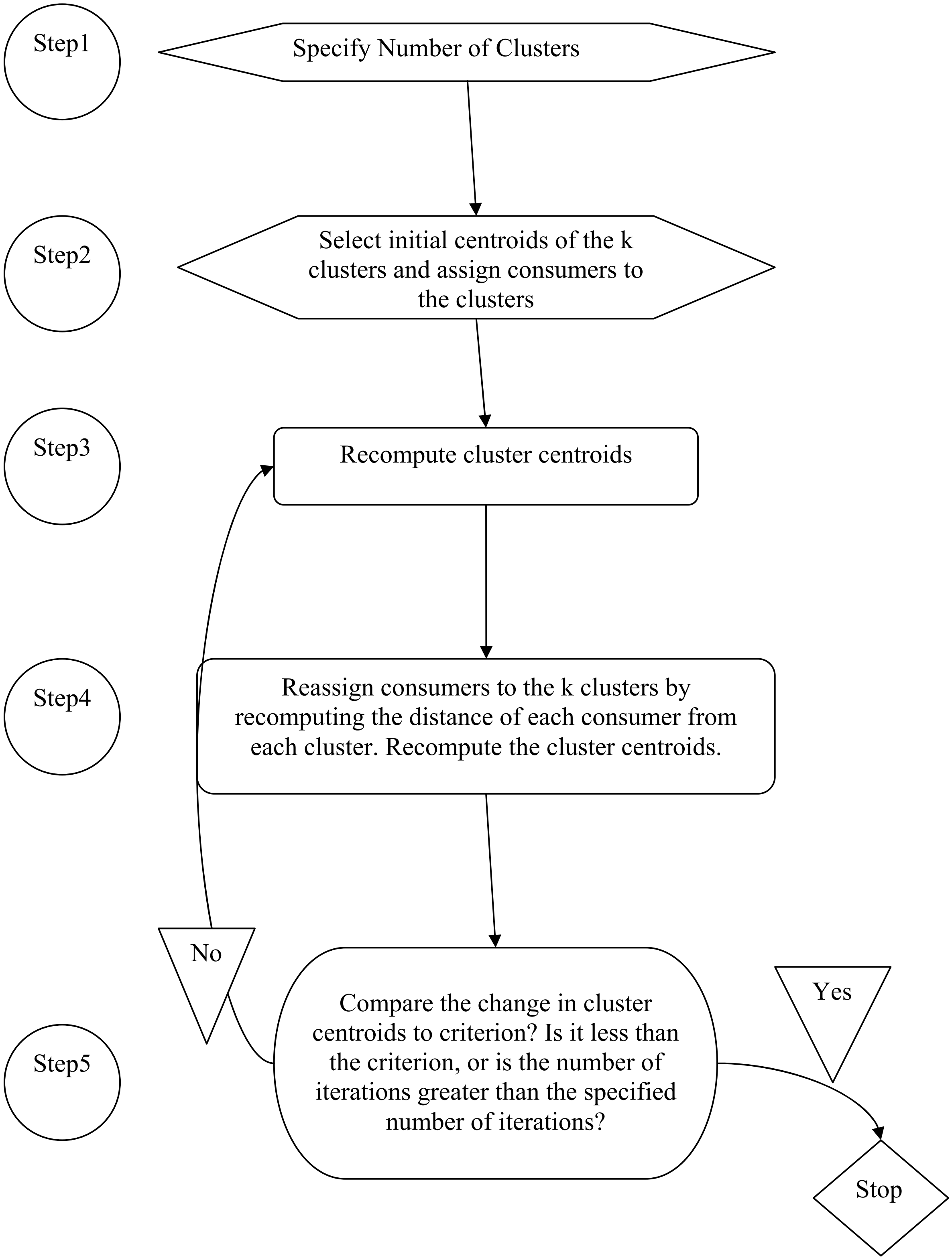


Figure 10.8

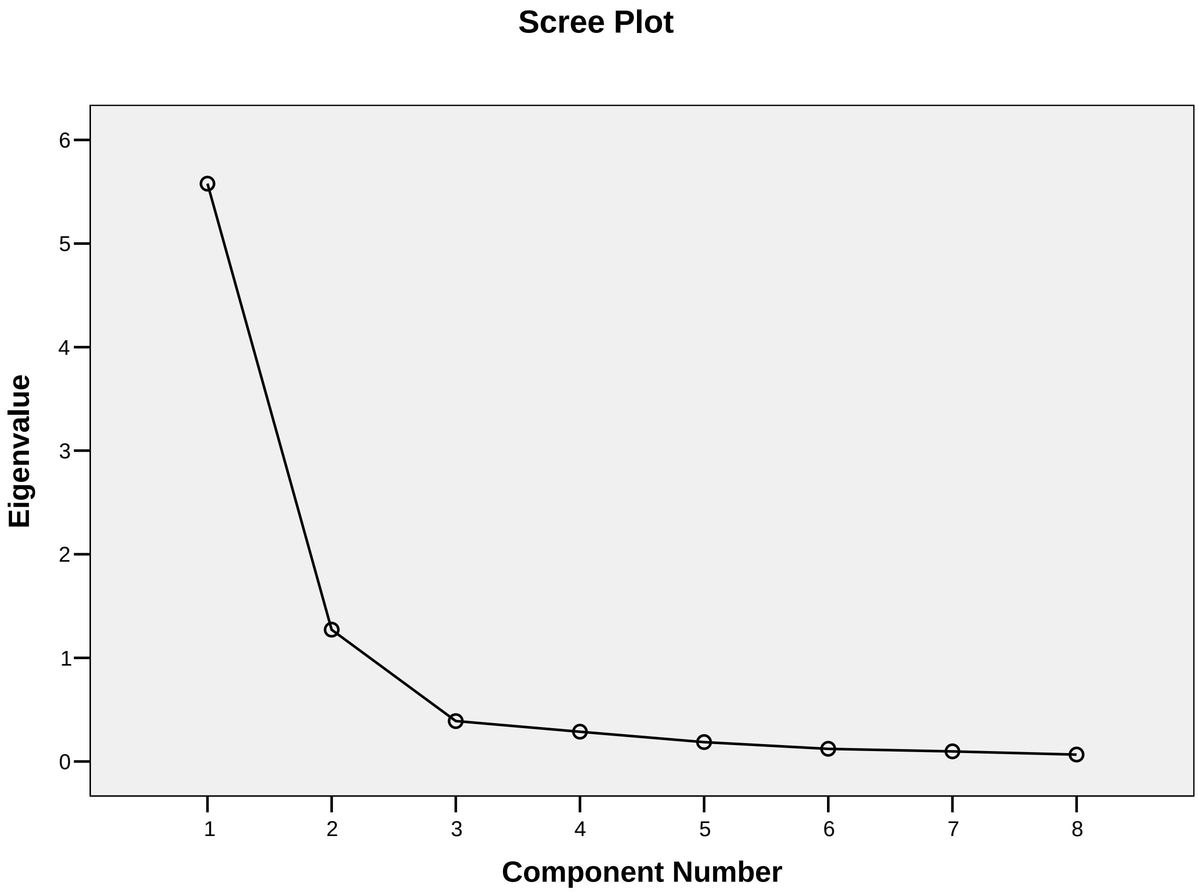
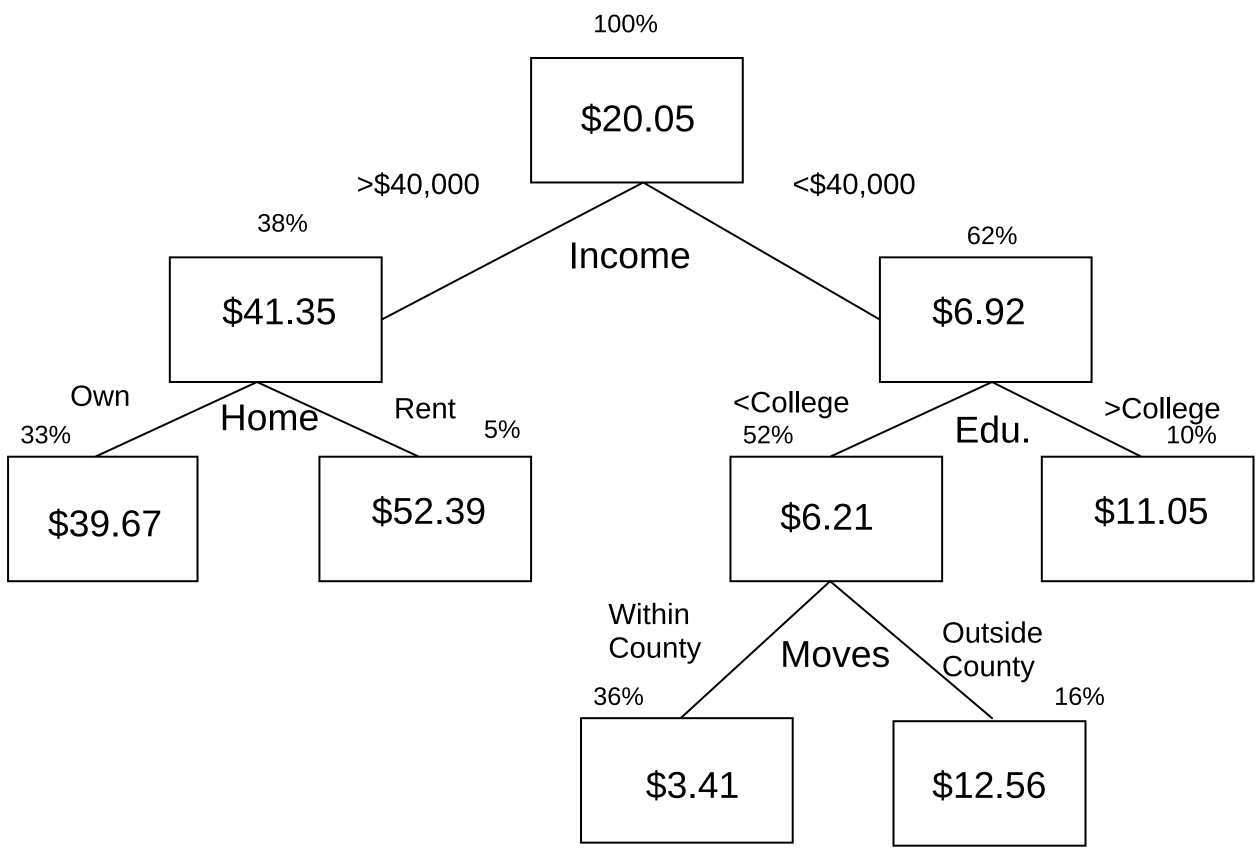


Figure 10.9



100%
$20.05
>$40,000
<$40,000
Income
38%
$41.35
62%
$6.92
Own
Home
Rent
33%
5%
$39.67
$52.39
<College
52%
Edu.
>College
10%
$6.21
$11.05
Within
County
Moves
Outside
County
36%
16%
$3.41
$12.56

[i] This example has been adapted from Sharma, Subash and Ajith Kumar, "Cluster Analysis and Factor Analysis," Handbook of Marketing Research: Uses, Misuses and Future Advances, Rajiv Grover and Marco Vriens, eds. Sage Publication, 2006

[ii] The following three books are reader-friendly references for many of the multivariate techniques discussed here. Sharma, S (1996), Applied Multivariate Techniques, New York: John Wiley. Hair, J., et.al.(2005), Multivariate Data Analysis, New York: Macmillan Publishing Company.

[iii] This example has been taken from the Latent Structure Regression chapter by Wayne, Kamakura, and Wedel in The Handbook of Marketing Research (2006), Grover and Vriens (eds.)

[iv] This example has been adapted from Dillon, William and Matthew Goldstein (1984), Multivariate Analysis: Methods and Applications, New York: John Wiley.

Concept Questions

1. What is the difference between the segmentation ideas discussed in Chapters1 and 3 and Data Driven Segmentation? When would you use one over the other?
2. What is the difference between bases of segmentation and profile of segments? Are they interchangeable from a mathematical point of view?
3. What is the difference between Hierarchical and Non-hierarchical Clustering? When would you use one over the other?
4. What is the difference between Single, Complete, Average and Centroid linkage methods? Can they give different solutions? Please illustrate.
5. How can different measurement scales influence a cluster solution?
6. What is the importance of the means (averages) of the variables in the final cluster solution?
7. What is Factor Analysis?
8. What are Factor Loadings and how are they used? What are Factor Score Coefficients and how are they used? What are Factor Scores?
9. What is the difference between Factor Analysis and Cluster Analysis?

10. Why would you use s combination of FA and CA?
11. What is Discriminant Analysis and when is it used?
12. What is the purpose of Latent Class Analysis? What value does it add over and above running regression on the entire sample?

Application Questions

Refer to the Needs, Attitudes and Behaviors Questionnaire in the Appendix C of Chapter 7. Design three empirical segmentation strategies using the three major segmentation techniques discussed in the chapter, i.e., Cluster Analysis, Factor/Cluster Analysis, and Latent Class Analysis. The content of the strategies would include what variables you would use as the basis of segmentation; what statistical techniques you would use; what variables you would use to profile these segments; and how would you profile the segments. Each strategy should use a different set of variables. Please discuss the insights each segmentation strategy would provide to the marketing manager and how might these insights be used to formulate marketing mix plans. Also hypothesis/discuss what kind of segments you might expect.