Cluster analysis

Business, Company

Chapter 9 Cluster Analysis Learning Objectives After reading this chapter you should understand: – The basic concepts of cluster analysis. – How basic cluster algorithms work. – How to compute simple clustering results manually. – The different types of clustering procedures. – The SPSS clustering outputs. Keywords Agglomerative and divisive clustering A Chebychev distance A City-block distance A Clustering variables A Dendrogram A Distance matrix A Euclidean distance A Hierarchical and partitioning methods A Icicle diagram A k-means A Matching coef? cients A Pro? ing clusters A Two-step clustering Are there any market segments where Web-enabled mobile telephony is taking off in different ways? To answer this question, Okazaki (2006) applies a twostep cluster analysis by identifying segments of Internet adopters in Japan. The ? ndings suggest that there are four clusters exhibiting distinct attitudes towards Web-enabled mobile telephony adoption. Interestingly, freelance, and highly educated professionals had the most negative perception of mobile Internet adoption, whereas clerical of? ce workers had the most positive perception.

Furthermore, housewives and company executives also exhibited a positive attitude toward mobile Internet usage. Marketing managers can now use these results to better target speci? c customer segments via mobile Internet services. Introduction Grouping similar customers and products is a fundamental marketing activity. It is used, prominently, in market segmentation. As companies cannot connect with all their customers, they have to divide markets into groups of consumers, customers, or clients (called segments) with similar needs and wants.

Firms can then target each of these segments by positioning themselves in a unique segment (such as Ferrari in the high-end sports car market). While market researchers often form E. Mooi and M. Sarstedt, A Concise Guide to Market Research, DOI 10. 1007/978-3-642-12541-6_9, # Springer-Verlag Berlin Heidelberg 2011 237 238 9 Cluster Analysis market segments based on practical grounds, industry practice and wisdom, cluster analysis allows segments to be formed that are based on data that are less dependent on subjectivity.

The segmentation of customers is a standard application of cluster analysis, but it can also be used in different, sometimes rather exotic, contexts such as evaluating typical supermarket shopping paths (Larson et al. 2005) or deriving employers' branding strategies (Moroko and Uncles 2009). Understanding Cluster Analysis Cluster analysis is a convenient method for identifying homogenous groups of objects called clusters. Objects (or cases, observations) in a speci? c cluster share many characteristics, but are very dissimilar to objects not belonging to that cluster.

Let's try to gain a basic understanding of the cluster analysis procedure by looking at a simple example. Imagine that you are interested in segmenting your customer base in order to better target them through, for example, pricing strategies. The ? rst step is to decide on the characteristics that you will use to segment your customers. In other words, you have to decide which clustering variables will be included in the analysis. For example, you may want to segment a market based on customers' price consciousness (x) and brandloyalty(y).

These two variables can be measured on a 7-point scale with higher values denoting a higher degree of price consciousness and brand loyalty. The values of seven respondents are shown in Table 9. 1 and the scatter plot in Fig. 9. 1. The objective of cluster analysis is to identify groups of objects (in this case, customers) that are very similar with regard to their price consciousness and brand loyalty and assign them into clusters. After having decided on the clustering variables (brand loyalty and price consciousness), we need to decide on the clustering procedure to form our groups of objects.

This step is crucial for the analysis, as different procedures require different decisions prior to analysis. There is an abundance of different approaches and little guidance on which one to use in practice. We are going to discuss the most popular approaches in market research, as they can be easily computed using SPSS. These approaches are: hierarchical methods, partitioning methods (more precisely, k-means), and two-step clustering, which is largely a combination of the ? rst two methods.

Each of these procedures follows a different approach to grouping the most similar objects into a cluster and to determining each object’s cluster membership. In other words, whereas an object in a certain cluster should be as similar as possible to all the other objects in the Table 9. 1 Data Customer x y A 3 7 B 6 7 C 5 6 D 3 5 E 6 5 F 4 3 G 1 2 Understanding Cluster Analysis 7 6 A C D E B 239 Brand loyalty (y) 5 4 3 2 1 0 0 1 2 G F 3 4 5 6 7 Price consciousness (x) Fig. 9. 1 Scatter plot same cluster, it should likewise be as distinct as possible from objects in different clusters. But how do we measure similarity?

Some approaches – most notably hierarchical methods – require us to specify how similar or different objects are in order to identify different clusters. Most software packages calculate a measure of (dis)similarity by estimating the distance between pairs of objects. Objects with smaller distances between one another are more similar, whereas objects with larger distances are more dissimilar. An important problem in the application of cluster analysis is the decision regarding how many clusters should be derived from the data. This question is explored in the next step of the analysis.

Sometimes, however, we already know the number of segments that have to be derived from the data. For example, if we were asked to ascertain what characteristics distinguish frequent shoppers from infrequent ones, we need to ? nd two different clusters. However, we do not usually know the exact number of clusters and then we face a trade-off. On the one hand, you want as few clusters as possible to make them easy to understand and actionable. On the other hand, having many clusters allows you to identify more segments and more subtle differences between segments.

In an extreme case, you can address each individual separately (called one-to-one marketing) to meet consumers' varying needs in the best possible way. Examples of such a micro-marketing strategy are Puma's Mongolian Shoe BBQ (www. mongolianshoebbq. puma. com) and Nike ID (http://nikeid. nike. com), in which customers can fully customize a pair of shoes in a hands-on, tactile, and interactive shoe-making experience. On the other hand, the costs associated with such a strategy may be prohibitively high in many 240 9 Cluster Analysis Decide on the clustering variables Decide on the clustering procedure

Hierarchical methods Select a measure of similarity or dissimilarity Partitioning methods Two-step clustering Select a measure of similarity or dissimilarity Choose a clustering algorithm Decide on the number of clusters Validate and interpret the cluster solution Fig. 9. 2 Steps in a cluster analysis business contexts. Thus, we have to ensure that the segments are large enough to make the targeted marketing programs pro? table. Consequently, we have to cope with a certain degree of within-cluster heterogeneity, which makes targeted marketing programs less effective.

In the ? nal step, we need to interpret the solution by de? ning and labeling the obtained clusters. This can be done by examining the clustering variables' mean values or by identifying explanatory variables to pro? le the clusters. Ultimately, managers should be able to identify customers in each segment on the basis of easily measurable variables. This ? nal step also requires us to assess the clustering solution's stability and validity. Figure 9. 2 illustrates the steps associated with a cluster analysis; we will discuss these in more detail in the following sections.

Conducting a Cluster Analysis Decide on the Clustering Variables At the beginning of the clustering process, we have to select appropriate variables for clustering. Even though this choice is of utmost importance, it is rarely treated as such and, instead, a mixture of intuition and data availability guide most analyses in marketing practice. However, faulty assumptions may lead to improper market Conducting a Cluster Analysis 241 segments and, consequently, to de? cient marketing strategies. Thus, great care should be taken when selecting the clustering variables. There are several types of clustering variables and these can be classi? d into general

(independent of products, services or circumstances) and speci? c (related to both the customer and the product, service and/or particular circumstance), on the one hand, and observable (i. e. , measured directly) and unobservable (i. e. , inferred) on the other. Table 9. 2 provides several types and examples of clustering variables. Table 9. 2 Types and examples of clustering variables General Observable (directly Cultural, geographic, demographic, measurable) socio-economic Unobservable Psychographics, values, personality, (inferred) lifestyle Adapted from Wedel and Kamakura (2000)

Speci? c User status, usage frequency, store and brand loyalty Bene? ts, perceptions, attitudes, intentions, preferences The types of variables used for cluster analysis provide different segments and, thereby, in? uence segment-targeting strategies. Over the last decades, attention has shifted from more traditional general clustering variables towards product-speci? c unobservable variables. The latter generally provide better guidance for decisions on marketing instruments' effective speci? cation. It is generally acknowledged that segments identi? ed by means of speci? unobservable variables are usually more homogenous and their consumers respond consistently to marketing actions (see Wedel and Kamakura 2000). However, consumers in these segments are also frequently hard to identify from variables that are easily measured, such as demographics. Conversely, segments determined by means of generally observable variables usually stand out due to their identi? ability but often lack a unique response structure. 1 Consequently, researchers often combine different variables (e. g. , multiple lifestyle characteristics combined with demographic variables), bene? ing from each ones strengths. In some cases, the choice of clustering

variables is apparent from the nature of the task at hand. For example, a managerial problem regarding corporate communications will have a fairly well de? ned set of clustering variables, including contenders such as awareness, attitudes, perceptions, and media habits. However, this is not always the case and researchers have to choose from a set of candidate variables. Whichever clustering variables are chosen, it is important to select those that provide a clear-cut differentiation between the segments regarding a speci? c managerial objective. More precisely, criterion validity is of special interest; that is, the extent to which the “ independent” clustering variables are associated with 1 2 See Wedel and Kamakura (2000). Tonks (2009) provides a discussion of segment design and the choice of clustering variables in consumer markets. 242 9 Cluster Analysis one or more “ dependent” variables not included in the analysis. Given this relationship, there should be signi? cant differences between the “ dependent” variable(s) across the clusters. These associations may or may not be causal, but it is essential that the clustering variables distinguish the “ dependent” variable(s) signi? antly. Criterion variables usually relate to some aspect of behavior, such as purchase intention or usage frequency. Generally, you should avoid using an abundance of clustering variables, as this increases the odds that the variables are no longer dissimilar. If there is a high degree of collinearity between the variables, they are not suf? ciently unique to identify distinct market segments. If highly correlated variables are used for cluster analysis, speci? c aspects covered by these variables will be overrepresented in the clustering solution.

In this regard, absolute correlations above 0. 90 are always problematic. For example, if we were to add another variable called brand preference to our analysis, it would virtually cover the same aspect as brand loyalty. Thus, the concept of being attached to a brand would be overrepresented in the analysis because the clustering procedure does not differentiate between the clustering variables in a conceptual sense. Researchers frequently handle this issue by applying cluster analysis to the observations' factor scores derived from a previously carried out factor analysis.

However, according to Dolnicar and Gr€n u (2009), this factor-cluster segmentation approach can lead to several problems: 1. The data are pre-processed and the clusters are identi? ed on the basis of transformed values, not on the original information, which leads to different results. 2. In factor analysis, the factor solution does not explain a certain amount of variance; thus, information is discarded before segments have been identi? ed or constructed. 3. Eliminating variables with low loadings on all the extracted factors means that, potentially, the most important pieces of information for the identi? ation of niche segments are discarded, making it impossible to ever identify such groups. 4. The interpretations of clusters based on the original variables become questionable given that the segments have been constructed using factor scores. Several studies have shown that the factor-cluster segmentation signi? cantly reduces the success of segment recovery. 3 Consequently, you should rather reduce the number of items in the questionnaire's pre-testing phase, retaining a reasonable number of relevant, non-redundant questions that you believe differentiate the segments well.

However, if you have your doubts about the data structure, factorclustering segmentation may still be a better option than discarding items that may conceptually be necessary. Furthermore, we should keep the sample size in mind. First and foremost, this relates to issues of managerial relevance as segments' sizes need to be substantial to ensure that targeted marketing programs are pro? table. From a statistical perspective, every additional variable requires an over-proportional increase in 3 See the studies by Arabie and Hubert (1994), Sheppard (1996), or Dolnicar and Gr€n (2009). u

Conducting a Cluster Analysis 243 observations to ensure valid results. Unfortunately, there is no generally accepted rule of thumb regarding minimum sample sizes or the relationship between the objects and the number of clustering variables used. In a related methodological context, Formann (1984) recommends a sample size of at least 2m, where m equals the number of clustering variables. This can only provide rough guidance; nevertheless, we should pay attention to the relationship between the objects and clustering variables. It does not, for example, appear logical to cluster ten objects using ten variables.

Keep in mind that no matter how many variables are used and no matter how small the sample size, cluster analysis will always render a result! Ultimately, the choice of clustering variables always depends on contextual in? uences such as data availability or resources to acquire additional data. Marketing researchers often overlook the fact that the choice of clustering variables is closely connected to data quality. Only those variables that ensure that high quality data can be used should be included in the analysis.

This is very important if a segmentation solution has to be managerially useful.

Furthermore, data are of high quality if the questions asked have a strong theoretical basis, are not contaminated by respondent fatigue or response styles, are recent, and thus re? ect the current market situation (Dolnicar and Lazarevski 2009). Lastly, the requirements of other managerial functions within the organization often play a major role. Sales and distribution may as well have a major in? uence on the design of market segments. Consequently, we have to be aware that subjectivity and common sense agreement will (and should) always impact the choice of clustering variables.

Decide on the Clustering Procedure By choosing a speci? c clustering procedure, we determine how clusters are to be formed. This always involves optimizing some kind of criterion, such as minimizing the within-cluster variance (i. e. , the clustering variables' overall variance of objects in a speci? c cluster), or maximizing the distance between the objects or clusters. The procedure could also address the question of how to determine the (dis)similarity between objects in a newly formed cluster and the remaining objects in the dataset.

There are many different clustering procedures and also many ways of classifying these (e. g. , overlapping versus non-overlapping, unimodal versus multimodal, exhaustive versus non-exhaustive). 4 A practical distinction is the differentiation between hierarchical and partitioning methods (most notably the k-means procedure), which we are going to discuss in the next sections. We also introduce two-step clustering, which

combines the principles of hierarchical and partitioning methods and which has recently gained increasing attention from market research practice.

See Wedel and Kamakura (2000), Dolnicar (2003), and Kaufman and Rousseeuw (2005) for a review of clustering techniques. 4 244 9 Cluster Analysis Hierarchical Methods Hierarchical clustering procedures are characterized by the tree-like structure established in the course of the analysis. Most hierarchical techniques fall into a category called agglomerative clustering. In this category, clusters are consecutively formed from objects. Initially, this type of procedure starts with each object representing an individual cluster.

These clusters are then sequentially merged according to their similarity. First, the two most similar clusters (i. e. , those with the smallest distance between them) are merged to form a new cluster at the bottom of the hierarchy. In the next step, another pair of clusters is merged and linked to a higher level of the hierarchy, and so on. This allows a hierarchy of clusters to be established from the bottom up. In Fig. 9. 3 (left-hand side), we show how agglomerative clustering assigns additional objects to clusters as the cluster size increases. Step 5 Step 1 A, B, C, D, E

Agglomerative clustering Step 4 Step 2 Divisive clustering A, B C, D, E Step 3 Step 3 A, B C, D E Step 2 Step 4 A, B C D E Step 1 Step 5 A B C D E Fig. 9. 3 Agglomerative and divisive clustering A cluster hierarchy can also be generated top-down. In this divisive clustering, all objects are initially merged into a single cluster, which is then gradually split up. Figure 9. 3 illustrates this concept (right-hand side). As we can see, in both

agglomerative and divisive clustering, a cluster on a higher level of the hierarchy always encompasses all clusters from a lower level.

This means that if an object is assigned to a certain cluster, there is no possibility of reassigning this object to another cluster. This is an important distinction between these types of clustering and partitioning methods such as k-means, which we will explore in the next section. Divisive procedures are quite rarely used in market research. We therefore concentrate on the agglomerative clustering procedures. There are various types Conducting a Cluster Analysis 245 of agglomerative procedures. However, before we discuss these, we need to de? ne how similarities or dissimilarities are measured between pairs of objects.

Select a Measure of Similarity or Dissimilarity There are various measures to express (dis)similarity between pairs of objects. A straightforward way to assess two objects' proximity is by drawing a straight line between them. For example, when we look at the scatter plot in Fig. 9. 1, we can easily see that the length of the line connecting observations B and C is much shorter than the line connecting B and G. This type of distance is also referred to as Euclidean distance (or straight-line distance) and is the most commonly used type when it comes to analyzing ratio or interval-scaled data. In our example, we have ordinal data, but market researchers usually treat ordinal data as metric data to calculate distance metrics by assuming that the scale steps are equidistant (very much like in factor analysis, which we discussed in Chap. 8). To use a hierarchical clustering procedure, we need to express these distances mathematically. By taking the data in Table 9. 1 into consideration, we can easily compute the Euclidean distance between

customer B and customer C (generally referred to as d(B, C)) with regard to the two variables x and y by using the following formula: q??? Euclidean ? B; C? ? ? xB A xC ? 2 ? ? yB A yC ? 2 The Euclidean distance is the square root of the sum of the squared differences in the variables' values. Using the data from Table 9. 1, we obtain the following: q??????????????????????????????????????? p??? dEuclidean ? B; C? ? ? 6 A 5? 2 ? ? 7 A 6? 2 ? 2 ? 1: 414 This distance corresponds to the length of the line that connects objects B and C. In this case, we only used two variables but we can easily add more under the root sign in the formula. However, each additional variable will add a dimension to our research problem (e. . , with six clustering variables, we have to deal with six dimensions), making it impossible to represent the solution graphically. Similarly, we can compute the distance between customer B and G, which yields the following: q??? p????? dEuclidean ? B; G? ? ? 6 A 1? 2 ? ? 7 A 2? 2 ? 50 ? 7: 071 Likewise, we can compute the distance between all other pairs of objects. All these distances are usually expressed by means of a distance matrix. In this distance matrix, the non-diagonal elements express the distances between pairs of objects 5

Note that researchers also often use the squared Euclidean distance. 246 9 Cluster Analysis and zeros on the diagonal (the distance from each object to itself is, of course, 0). In our example, the distance matrix is an 8 A 8 table with the lines and rows representing the objects (i. e. , customers) under consideration (see Table 9. 3). As the distance between objects B and C (in this case 1. 414 units) is the same as between C and B, the distance matrix

is symmetrical. Furthermore, since the distance between an object and itself is zero, one need only look at either the lower or upper non-diagonal elements.

Table 9. 3 Euclidean distance matrix Objects A B A 0 B 3 0 C 2. 236 1. 414 D 2 3. 606 E 3. 606 2 F 4. 123 4. 472 G 5. 385 7. 071 C D E F G 0 2. 236 1. 414 3. 162 5. 657 0 3 2. 236 3. 606 0 2. 828 5. 831 0 3. 162 0 There are also alternative distance measures: The city-block distance uses the sum of the variables' absolute differences. This is often called the Manhattan metric as it is akin to the walking distance between two points in a city like New York's Manhattan district, where the distance equals the number of blocks in the directions North-South and East-West.

Using the city-block distance to compute the distance between customers B and C (or C and B) yields the following: dCityAblock ? B; C? ? jxB A xC j ? jyB A yC j ? j6 A 5j ? j7 A 6j ? 2 The resulting distance matrix is in Table 9. 4. Table 9. 4 City-block distance matrix Objects A B A 0 B 3 0 C 3 2 D 2 5 E 5 2 F 5 6 G 7 10 C D E F G 0 3 2 4 8 0 3 3 5 0 4 8 0 4 0 Lastly, when working with metric (or ordinal) data, researchers frequently use the Chebychev distance, which is the maximum of the absolute difference in the clustering variables' values. Inrespectof customers B and C, this result is: dChebychec ? B; C? max? jxB A xC j; jyB A yC j? ? max? j6 A 5j; j7 A 6j? ? 1 Figure 9. 4 illustrates the interrelation between these three distance measures regarding two objects, C and G, from our example. Conducting a Cluster Analysis 247 C Brand loyalty (y) Euclidean distance City-block distance G Chebychev distance Price consciousness (x) Fig. 9. 4 Distance measures There are other distance measures such as the Angular, Canberra or Mahalanobis distance.

In many situations, the latter is desirable as it compensates for collinearity between the clustering variables. However, it is (unfortunately) not menu-accessible in SPSS.

In many analysis tasks, the variables under consideration are measured on different scales or levels. This would be the case if we extended our set of clustering variables by adding another ordinal variable representing the customers' income measured by means of, for example, 15 categories. Since the absolute variation of the income variable would be much greater than the variation of the remaining two variables (remember, that x and y are measured on 7-point scales), this would clearly distort our analysis results. We can resolve this problem by standardizing the data prior to the analysis.

Different standardization methods are available, such as the simple z standardization, which rescales each variable to have a mean of 0 and a standard deviation of 1 (see Chap. 5). In most situations, however, standardization by range (e. g. , to a range of 0 to 1 or A1 to 1) performs better. 6 We recommend standardizing the data in general, even though this procedure can reduce or in? ate the variables' in? uence on the clustering solution. 6 See Milligan and Cooper (1988). 248 9 Cluster Analysis Another way of (implicitly) standardizing the data is by using the correlation between the objects instead of distance measures.

For example, suppose a respondent rated price consciousness 2 and brand loyalty 3. Now suppose a second respondent indicated 5 and 6, whereas a third rated these variables 3 and 3. Euclidean, city-block, and Chebychev distances would indicate that the ? rst respondent is more similar to the third than to the second. Nevertheless, one could convincingly argue that the ? rst

respondent’s ratings are more similar to the second’s, as both rate brand loyalty higher than price consciousness. This can be accounted for by computing the correlation between two vectors of values as a measure of similarity (i. . , high correlation coef? cients indicate a high degree of similarity). Consequently, similarity is no longer de? ned by means of the difference between the answer categories but by means of the similarity of the answering pro? les. Using correlation is also a way of standardizing the data implicitly. Whether you use correlation or one of the distance measures depends on whether you think the relative magnitude of the variables within an object (which favors correlation) matters more than the relative magnitude of each variable across objects (which favors distance).

However, it is generally recommended that one uses correlations when applying clustering procedures that are susceptible to outliers, such as complete linkage, average linkage or centroid (see next section). Whereas the distance measures presented thus far can be used for metrically and – in general – ordinally scaled data, applying them to nominal or binary data is meaningless. In this type of analysis, you should rather select a similarity measure expressing the degree to which variables’ values share the same category. These socalled matching coef? ients can take different forms but rely on the same allocation scheme shown in Table 9. 5. Table 9. 5 Allocation scheme for matching coef? cients Number of variables with category 1 a c Object 1 Number of variables with category 2 b d Object 2 Number of variables with category 1 Number of variables with category 2 Based on the allocation scheme in Table 9. 5, we can compute different matching coef? cients, such as the simple matching coef? cient (SM): SM ? a? d a? b? c? d

This coef? cient is useful when both positive and negative values carry an equal degree of information.

For example, gender is a symmetrical attribute because the number of males and females provides an equal degree of information. Conducting a Cluster Analysis 249 Let’s take a look at an example by assuming that we have a dataset with three binary variables: gender (male ? 1, female ? 2), customer (customer ? 1, noncustomer ? 2), and disposable income (low ? 1, high ? 2). The ? rst object is a male non-customer with a high disposable income, whereas the second object is a female non-customer with a high disposable income. According to the scheme in Table 9. , a ? b ? 0, c ? 1 and d ? 2, with the simple matching coef? cient taking a value of 0. 667. Two other types of matching coef? cients, which do not equate the joint absence of a characteristic with similarity and may, therefore, be of more value in segmentation studies, are the Jaccard (JC) and the Russel and Rao (RR) coef? cients. They are de? ned as follows: a JC ? a? b? c a RR ? a? b? c? d These matching coef? cients are – just like the distance measures – used to determine a cluster solution. There are many other matching coef? ients such as Yule’s Q, Kulczynski or Ochiai, but since most applications of cluster analysis rely on metric or ordinal data, we will not discuss these in greater detail. 7 For nominal variables with more than two categories, you should always convert the categorical variable into a set of binary variables in order to use matching coef? cients. When you have ordinal data, you should always use distance measures such as Euclidean distance. Even though using matching coef? cients would be feasible and – from a strictly statistical

standpoint – even more appropriate, you would disregard variable information in the sequence of the categories.

In the end, a respondent who indicates that he or she is very loyal to a brand is going to be closer to someone who is somewhat loyal than a respondent who is not loyal at all. Furthermore, distance measures best represent the concept of proximity, which is fundamental to cluster analysis. Most datasets contain variables that are measured on multiple scales. For example, a market research questionnaire may ask about the respondent's income, product ratings, and last brand purchased. Thus, we have to consider variables measured on a ratio, ordinal, and nominal scale. How can we simultaneously incorporate these variables into one analysis?

Unfortunately, this problem cannot be easily resolved and, in fact, many market researchers simply ignore the scale level. Instead, they use one of the distance measures discussed in the context of metric (and ordinal) data. Even though this approach may slightly change the results when compared to those using matching coef? cients, it should not be rejected. Cluster analysis is mostly an exploratory technique whose results provide a rough guidance for managerial decisions. Despite this, there are several procedures that allow a simultaneous integration of these variables into one analysis. 7

See Wedel and Kamakura (2000) for more information on alternative matching coef? cients. 250 9 Cluster Analysis First, we could compute distinct distance matrices for each group of variables; that is, one distance matrix based on, for example, ordinally scaled variables and another based on nominal variables. Afterwards, we can simply compute the weighted

arithmetic mean of the distances and use this average distance matrix as the input for the cluster analysis. However, the weights have to be determined a priori and improper weights may result in a biased treatment of different variable types.

Furthermore, the computation and handling of distance matrices are not trivial. Using the SPSS syntax, one has to manually add the MATRIX subcommand, which exports the initial distance matrix into a new data ? le. Go to the 8 Web Appendix (! Chap. 5) to learn how to modify the SPSS syntax accordingly. Second, we could dichotomize all variables and apply the matching coef? cients discussed above. In the case of metric variables, this would involve specifying categories (e. g. , low, medium, and high income) and converting these into sets of binary variables. In most cases, however, the speci? ation of categories would be rather arbitrary and, as mentioned earlier, this procedure could lead to a severe loss of information. In the light of these issues, you should avoid combining metric and nominal variables in a single cluster analysis, but if this is not feasible, the two-step clustering procedure provides a valuable alternative, which we will discuss later. Lastly, the choice of the (dis)similarity measure is not extremely critical to recovering the underlying cluster structure. In this regard, the choice of the clustering algorithm is far more important.

We therefore deal with this aspect in the following section. Select a Clustering Algorithm After having chosen the distance or similarity measure, we need to decide which clustering algorithm to apply. There are several agglomerative procedures and they can be distinguished by the way they de? ne the distance from a newly formed cluster to a certain object, or to

other clusters in the solution. The most popular agglomerative clustering procedures include the following: l l l l Single linkage (nearest neighbor): The distance between two clusters corresponds to the shortest distance between any two members in the two clusters.

Complete linkage (furthest neighbor): The oppositional approach to single linkage assumes that the distance between two clusters is based on the longest distance between any two members in the two clusters. Average linkage: The distance between two clusters is de? ned as the average distance between all pairs of the two clusters' members. Centroid: In this approach, the geometric center (centroid) of each cluster is computed ? rst. The distance between the two clusters equals the distance between the two centroids. Figures 9. 5–9. 8 illustrate these linkage procedures for two randomly framed clusters.

Conducting a Cluster Analysis Fig. 9. 5 Single linkage 251 Fig. 9. 6 Complete linkage Fig. 9. 7 Average linkage Fig. 9. 8 Centroid 252 9 Cluster Analysis Each of these linkage algorithms can yield totally different results when used on the same dataset, as each has its speci? c properties. As the single linkage algorithm is based on minimum distances, it tends to form one large cluster with the other clusters containing only one or few objects each. We can make use of this " chaining effect" to detect outliers, as these will be merged with the remaining objects – usually at very large distances – in the last steps of the analysis.

Generally, single linkage is considered the most versatile algorithm. Conversely, the complete linkage method is strongly affected by outliers, as it is based on maximum distances. Clusters produced by this method are

likely to be rather compact and tightly clustered. The average linkage and centroid algorithms tend to produce clusters with rather low within-cluster variance and similar sizes. However, both procedures are affected by outliers, though not as much as complete linkage. Another commonly used approach in hierarchical clustering is Ward's method. This approach does not combine the two most similar objects successively.

Instead, those objects whose merger increases the overall within-cluster variance to the smallest possible degree, are combined. If you expect somewhat equally sized clusters and the dataset does not include outliers, you should always use Ward's method. To better understand how a clustering algorithm works, let's manually examine some of the single linkage procedure's calculation steps. We start off by looking at the initial (Euclidean) distance matrix in Table 9. 3. In the very ? rst step, the two objects exhibiting the smallest distance in the matrix are merged.

Note that we always merge those objects with the smallest distance, regardless of the clustering procedure (e. g. , single or complete linkage). As we can see, this happens to two pairs of objects, namely B and C (d(B, C) ? 1. 414), as well as C and E (d(C, E) ? 1. 414). In the next step, we will see that it does not make any difference whether we ? rst merge the one or the other, so let's proceed by forming a new cluster, using objects B and C. Having made this decision, we then form a new distance matrix by considering the single linkage decision rule as discussed above.

According to this rule, the distance from, for example, object A to the newly formed cluster is the minimum of d(A, B) and d(A, C). As d(A, C) is smaller than d(A, B), the distance from A to the newly formed cluster is equal to d(A,

C); that is, 2. 236. We also compute the distances from cluster [B, C] (clusters are indicated by means of squared brackets) to all other objects (i. e. D, E, F, G) and simply copy the remaining distances – such as d(E, F) – that the previous clustering has not affected. This yields the distance matrix shown in Table 9. 6.

Continuing the clustering procedure, we simply repeat the last step by merging the objects in the new distance matrix that exhibit the smallest distance (in this case, the newly formed cluster [B, C] and object E) and calculate the distance from this cluster to all other objects. The result of this step is described in Table 9. 7. Try to calculate the remaining steps yourself and compare your solution with the distance matrices in the following Tables 9. 8–9. 10. Conducting a Cluster Analysis Table 9. 6 Distance matrix after ? rst clustering step (single linkage) Objects A B, C D E F G A 0 B, C 2. 36 0 D 2 2. 236 0 E 3. 606 1. 414 3 0 F 4. 123 3. 162 2. 236 2. 828 0 G 5. 385 5. 657 3. 606 5. 831 3. 162 0 253 Table 9. 7 Distance matrix after second clustering step (single linkage) Objects A B, C, E D F G A 0 B, C, E 2. 236 0 D 2 2. 236 0 F 4. 123 2. 828 2. 236 0 G 5. 385 5. 657 3. 606 3. 162 0 Table 9. 8 Distance matrix after third clustering step (single linkage) Objects A, D B, C, E F G A, D 0 B, C, E 2. 236 0 F 2. 236 2. 828 0 G 3. 606 5. 657 3. 162 0 Table 9. 9 Distance matrix after fourth clustering step (single linkage) Objects A, B, C, D, E F G A, B, C, D, E 0 F 2. 236 0 G 3. 06 3. 162 0 Table 9. 10 Distance matrix after ? fth clustering step (single linkage) Objects A, B, C, D, E, F G A, B, C, D, E, F 0 G 3. 162 0 By following the single linkage procedure, the last steps involve the merger of cluster [A, B, C, D, E, F] and object G at a distance of 3. 162. Do you get the same results? As you can see, conducting

a basic cluster analysis manually is not that hard at all – not if there are only a few objects in the dataset. A common way to visualize the cluster analysis's progress is by drawing a dendrogram, which displays the distance level at which there was a ombination of objects and clusters (Fig. 9. 9). We read the dendrogram from left to right to see at which distance objects have been combined. For example, according to our calculations above, objects B, C, and E are combined at a distance level of 1. 414. 254 B C E A D F G 9 Cluster Analysis 0 1 2 Distance 3 Fig. 9. 9 Dendrogram Decide on the Number of Clusters An important question we haven't yet addressed is how to decide on the number of clusters to retain from the data. Unfortunately, hierarchical methods provide only very limited guidance for making this decision.

The only meaningful indicator relates to the distances at which the objects are combined. Similar to factor analysis's scree plot, we can seek a solution in which an additional combination of clusters or objects would occur at a greatly increased distance. This raises the issue of what a great distance is, of course. One potential way to solve this problem is to plot the number of clusters on the x-axis (starting with the one-cluster solution at the very left) against the distance at which objects or clusters are combined on the y-axis.

Using this plot, we then search for the distinctive break (elbow). SPSS does not produce this plot automatically – you have to use the distances provided by SPSS to draw a line chart by using a common spreadsheet program such as Microsoft Excel. Alternatively, we can make use of the dendrogram which essentially carries the same information. SPSS provides a dendrogram; however, this differs slightly from the one presented in Fig. 9. 9. Speci? cally,

SPSS rescales the distances to a range of 0–25; that is, the last merging step to a one-cluster solution takes place at a (rescaled) distance of 25.

The rescaling often lengthens the merging steps, thus making breaks occurring at a greatly increased distance level more obvious. Despite this, this distance-based decision rule does not work very well in all cases. It is often dif? cult to identify where the break actually occurs. This is also the case in our example above. By looking at the dendrogram, we could justify a two-cluster solution ([A, B, C, D, E, F] and [G]), as well as a ? ve-cluster solution ([B, C, E], [A], [D], [F], [G]). Conducting a Cluster Analysis 255 Research has suggested several other procedures for determining the number of clusters in a dataset.

Most notably, the variance ratio criterion (VRC) by Calinski and Harabasz (1974) has proven to work well in many situations. 8 For a solution with n objects and k segments, the criterion is given by: VRCk ? ? SSB =? k A 1?? =? SSW =? n A k?? ; where SSB is the sum of the squares between the segments and SSW is the sum of the squares within the segments. The criterion should seem familiar, as this is nothing but the F-value of a one-way ANOVA, with k representing the factor levels. Consequently, the VRC can easily be computed using SPSS, even though it is not readily available in the clustering procedures' outputs.

To ? nally determine the appropriate number of segments, we compute ok for each segment solution as follows: ok ? ? VRCk? 1 A VRCk ? A ? VRCk A VRCkA1 ? : In the next step, we choose the number of segments k that minimizes the value in ok. Owing to the term VRCkA1, the minimum number of clusters that can be selected is three, which is a clear disadvantage of the

criterion, thus limiting its application in practice. Overall, the data can often only provide rough guidance regarding the number of clusters you should select; consequently, you should rather revert to practical considerations.

Occasionally, you might have a priori knowledge, or a theory on which you can base your choice. However, ? rst and foremost, you should ensure that your results are interpretable and meaningful. Not only must the number of clusters be small enough to ensure manageability, but each segment should also be large enough to warrant strategic attention. Partitioning Methods: k-means Another important group of clustering procedures are partitioning methods. As with hierarchical clustering, there is a wide array of different algorithms; of these, the k-means procedure is the most important one for market research. The k-means algorithm follows an entirely different concept than the hierarchical methods discussed before. This algorithm is not based on distance measures such as Euclidean distance or city-block distance, but uses the within-cluster variation as a Milligan and Cooper (1985) compare various criteria. Note that the k-means algorithm is one of the simplest non-hierarchical clustering methods. Several extensions, such as k-medoids (Kaufman and Rousseeuw 2005) have been proposed to handle problematic aspects of the procedure. More advanced methods include ? ite mixture models (McLachlan and Peel 2000), neural networks (Bishop 2006), and self-organizing maps (Kohonen 1982). Andrews and Currim (2003) discuss the validity of some of these approaches. 9 8 256 9 Cluster Analysis measure to form homogenous clusters. Speci? cally, the procedure aims at segmenting the data in such a way that the within-cluster variation is minimized. Consequently, we do not need to decide on a distance measure in the ? rst

step of the analysis. The clustering process starts by randomly assigning objects to a number of clusters. 0 The objects are then successively reassigned to other clusters to minimize the within-cluster variation, which is basically the (squared) distance from eachobservationto the center of the associated cluster. If the reallocation of an object to another cluster decreases the within-cluster variation, this object is reassigned to that cluster. With the hierarchical methods, an object remains in a cluster once it is assigned to it, but with k-means, cluster af? liations can change in the course of the clustering process. Consequently, k-means does not build a hierarchy as described before (Fig. . 3), which is why the approach is also frequently labeled as non-hierarchical. For a better understanding of the approach, let’s take a look at how it works in practice. Figs. 9. 10–9. 13 illustrate the k-means clustering process. Prior to analysis, we have to decide on the number of clusters. Our client could, for example, tell us how many segments are needed, or we may know from previous research what to look for. Based on this information, the algorithm randomly selects a center for each cluster (step 1). In our example, two cluster centers are randomly initiated, which CC1 (? st cluster) and CC2 (second cluster) in Fig. 9. 10 A CC1 C B D E Brand loyalty (y) CC2 F G Price consciousness (x) Fig. 9. 10 k-means procedure (step 1) 10 Note this holds for the algorithms original design. SPSS does not choose centers randomly. Conducting a Cluster Analysis A CC1 C B 257 D E Brand loyalty (y) CC2 F G Price consciousness (x) Fig. 9. 11 k-means procedure (step 2) A CC1 CC1? C B Brand loyalty (y) D E CC2 CC2? F G Price consciousness (x) Fig. 9. 12 k-means procedure (step 3) 258 A CC1? 9 Cluster Analysis B C Brand loyalty (y) D E CC2? F G Price

consciousness (x) Fig. 9. 13 k-means procedure (step 4) epresent. 11 After this (step 2), Euclidean distances are computed from the cluster centers to every single object. Each object is then assigned to the cluster center with the shortest distance to it. In our example (Fig. 9. 11), objects A, B, and C are assigned to the ? rst cluster, whereas objects D, E, F, and G are assigned to the second. We now have our initial partitioning of the objects into two clusters. Based on this initial partition, each cluster's geometric center (i. e. , its centroid) is computed (third step). This is done by computing the mean values of the objects contained in the cluster (e. . , A, B, C in the ? rst cluster) regarding each of the variables (price consciousness and brand loyalty). As we can see in Fig. 9. 12, both clusters' centers now shift into new positions (CC1' for the ? rst and CC2' for the second cluster). In the fourth step, the distances from each object to the newly located cluster centers are computed and objects are again assigned to a certain cluster on the basis of their minimum distance to other cluster centers (CC1' and CC2'). Since the cluster centers' position changed with respect to the initial situation in the ? st step, this could lead to a different cluster solution. This is also true of our example, as object E is now – unlike in the initial partition – closer to the ? rst cluster center (CC1') than to the second (CC2'). Consequently, this object is now assigned to the ? rst cluster (Fig. 9. 13). The k-means procedure now repeats the third step and re-computes the cluster centers of the newly formed clusters, and so on. In other 11 Conversely, SPSS always sets one observation as the cluster center instead of picking some random point in the dataset. Conducting a Cluster Analysis 59 words, steps 3 and 4 are repeated until a predetermined number of iterations are reached, or

convergence is achieved (i. e. , there is no change in the cluster af? liations). Generally, k-means is superior to hierarchical methods as it is less affected by outliers and the presence of irrelevant clustering variables. Furthermore, k-means can be applied to very large datasets, as the procedure is less computationally demanding than hierarchical methods. In fact, we suggest de? nitely using k-means for sample sizes above 500, especially if many clustering variables are used.

From a strictly statistical viewpoint, k-means should only be used on interval or ratioscaled data as the procedure relies on Euclidean distances. However, the procedure is routinely used on ordinal data as well, even though there might be some distortions. One problem associated with the application of k-means relates to the fact that the researcher has to pre-specify the number of clusters to retain from the data. This makes k-means less attractive to some and still hinders its routine application in practice. However, the VRC discussed above can likewise be used for k-means clustering an application of this index can be found in the 8 Web Appendix ! Chap. 9). Another workaround that many market researchers routinely use is to apply a hierarchical procedure to determine the number of clusters and k-means afterwards. 12 This also enables the user to ? nd starting values for the initial cluster centers to handle a second problem, which relates to the procedure's sensitivity to the initial classi? cation (we will follow this approach in the example application). Two-Step Clustering We have already discussed the issue of analyzing mixed variables measured on different scale levels in this chapter.

The two-step cluster analysis developed by Chiu et al. (2001) has been speci? cally designed to handle this problem. Like k-means, the procedure can also effectively cope with very large datasets. The name two-step clustering is already an indication that the algorithm is based on a two-stage approach: In the ? rst stage, the algorithm undertakes a procedure that is very similar to the k-means algorithm. Based on these results, the two-step procedure conducts a modi? ed hierarchical agglomerative clustering procedure that combines the objects sequentially to form homogenous clusters.

This is done by building a so-called cluster feature tree whose “ leaves” represent distinct objects in the dataset. The procedure can handle categorical and continuous variables simultaneously and offers the user the ? exibility to specify the cluster numbers as well as the maximum number of clusters, or to allow the technique to automatically choose the number of clusters on the basis of statistical evaluation criteria. Likewise, the procedure guides the decision of how many clusters to retain from the data by calculating measures-of-? t such as Akaike’s Information Criterion (AIC) or Bayes 2 See Punji and Stewart (1983) for additional information on this sequential approach. 260 9 Cluster Analysis Information Criterion (BIC). Furthermore, the procedure indicates each variable’s importance for the construction of a speci? c cluster. These desirable features make the somewhat less popular two-step clustering a viable alternative to the traditional methods. You can ? nd a more detailed discussion of the two-step clustering procedure in the 8 Web Appendix (! Chap. 9), but we will also apply this method in the subsequent example.

Validate and Interpret the Cluster Solution Before interpreting the cluster solution, we have to assess the solution's stability and validity. Stability is evaluated by using different clustering procedures on the same data and testing whether these yield the same results. In hierarchical clustering, you can likewise use different distance measures. However, please note that it is common for results to change even when your solution is adequate. How much variation you should allow before questioning the stability of your solution is a matter of taste.

Another common approach is to split the dataset into two halves and to thereafter analyze the two subsets separately using the same parameter settings. You then compare the two solutions' cluster centroids. If these do not differ signi? cantly, you can presume that the overall solution has a high degree of stability. When using hierarchical clustering, it is also worthwhile changing the order of the objects in your dataset and re-running the analysis to check the results' stability. The results should not, of course, depend on the order of the dataset. If they do, you should try to ascertain if any obvious outliers may in? ence the results of the change in order. Assessing the solution's reliability is closely related to the above, as reliability refers to the degree to which the solution is stable over time. If segments quickly change their composition, or its members their behavior, targeting strategies are likely not to succeed. Therefore, a certain degree of stability is necessary to ensure that marketing strategies can be implemented and produce adequate results. This can be evaluated by critically revisiting and replicating the clustering results at a later point in time. To validate the clustering solution, we need to assess its criterion validity.

In research, we could focus on criterion variables that have a theoretically based relationship with the clustering variables, but were not included in the analysis. In market research, criterion variables usually relate to managerial outcomes such as the sales per person, or satisfaction. If these criterion variables differ signi? cantly, we can conclude that the clusters are distinct groups with criterion validity. To judge validity, you should also assess face validity and, if possible, expert validity. While we primarily consider criterion validity when choosing clustering variables, as well as in this ? al step of the analysis procedure, the assessment of face validity is a process rather than a single event. The key to successful segmentation is to critically revisit the results of different cluster analysis set-ups (e. g. , by using Conducting a Cluster Analysis 261 different algorithms on the same data) in terms of managerial relevance. This underlines the exploratory character of the method. The following criteria will help you make an evaluation choice for a clustering solution (Dibb 1999; Tonks 2009; Kotler and Keller 2009). l l l l l l l l l Substantial: The segments are large and pro? able enough to serve. Accessible: The segments can be effectively reached and served, which requires them to be characterized by means of observable variables. Differentiable: The segments can be distinguished conceptually and respond differently to different marketing-mix elements and programs. Actionable: Effective programs can be formulated to attract and serve the segments. Stable: Only segments that are stable over time can provide the necessary grounds for a successful marketing strategy. Parsimonious: To be managerially meaningful, only a small set of substantial clusters should be identi? ed.

Familiar: To ensure management acceptance, the segments composition should be comprehensible. Relevant: Segments should be relevant in respect of the company's competencies and objectives. Compactness: Segments exhibit a high degree of within-segment homogeneity and between-segment heterogeneity. Compatibility: Segmentation results meet other managerial functions' requirements. The ? nal step of any cluster analysis is the interpretation of the clusters. Interpreting clusters always involves examining the cluster centroids, which are the clustering variables' average values of all objects in a certain cluster.

This step is of the utmost importance, as the analysis sheds light on whether the segments are conceptually distinguishable. Only if certain clusters exhibit signi? cantly different means in these variables are they distinguishable – from a data perspective, at least. This can easily be ascertained by comparing the clusters with independent t-tests samples or ANOVA (see Chap. 6). By using this information, we can also try to come up with a meaningful name or label for each cluster; that is, one which adequately re? ects the objects in the cluster.

This is usually a very challenging task. Furthermore, clustering variables are frequently unobservable, which poses another problem. How can we decide to which segment a new object should be assigned if its unobservable characteristics, such as personality traits, personal values or lifestyles, are unknown? We could obviously try to survey these attributes and make a decision based on the clustering variables. However, this will not be feasible in most situations and researchers therefore try to identify observable variables that best mirror the partition of the objects.

If it is possible to identify, for example, demographic variables leading to a very similar partition as that obtained through the segmentation, then it is easy to assign a new object to a certain segment on the basis of these demographic 262 9 Cluster Analysis characteristics. These variables can then also be used to characterize speci? c segments, an action commonly called pro? ling. For example, imagine that we used a set of items to assess the respondents' values and learned that a certain segment comprises respondents who appreciate self-ful? lment, enjoyment of life, and a sense of accomplishment, whereas this is not the case in another segment. If we were able to identify explanatory variables such as gender or age, which adequately distinguish these segments, then we could partition a new person based on the modalities of these observable variables whose traits may still be unknown. Table 9. 11 summarizes the steps involved in a hierarchical and k-means clustering. While companies often develop their own market segments, they frequently use standardized segments, which are based on established buying trends, habits, and customers' needs and have been speci? ally designed for use by many products in mature markets. One of the most popular approaches is the PRIZM lifestyle segmentation system developed by Claritas Inc. , a leading market research company. PRIZM de? nes every US household in terms of 66 demographically and behaviorally distinct segments to help marketers discern those consumers' likes, dislikes, lifestyles, and purchase behaviors. Visit the Claritas website and ? ip through the various segment pro? les. By entering a 5-digit US ZIP code, you can also ? nd a speci? c neighborhood's top ? ve lifestyle groups.

One example of a segment is “ Gray Power,” containing middle-class, homeowning suburbanites who are aging in place rather than moving to retirement communities. Gray Power re? ects this trend, a segment of older, midscale singles and couples who live in quiet comfort. http://www. claritas. com/MyBestSegments/Default. jsp We also introduce steps related to two-step clustering which we will further introduce in the subsequent example. Conducting a Cluster Analysis 263 Table 9. 11 Steps involved in carrying out a factor analysis in SPSS Theory Action Research problem Identi? ation of homogenous groups of objects in a population Select clustering variables that should be Select relevant variables that potentially exhibit used to form segments high degrees of criterion validity with regard to a speci? c managerial objective. Requirements Suf? cient sample size Make sure that the relationship between objects and clustering variables is reasonable (rough guideline: number of observations should be at least 2m, where m is the number of clustering variables). Ensure that the sample size is large enough to guarantee substantial segments. Low levels of collinearity among the variables ?

Analyze ? Correlate ? Bivariate Eliminate or replace highly correlated variables (correlation coef? cients > 0. 90). Speci? cation Choose the clustering procedure If there is a limited number of objects in your dataset or you do not know the number of clusters: ? Analyze ? Classify ? Hierarchical Cluster If there are many observations (> 500) in your dataset and you have a priori knowledge regarding the number of clusters: ? Analyze ? Classify ? K-Means Cluster If there are many observations in your dataset and the

clustering variables are measured on different scale levels: ? Analyze ? Classify ?

Two-Step Cluster Select a measure of similarity or dissimilarity Hierarchical methods: (only hierarchical and two-step clustering) ? Analyze ? Classify ? Hierarchical Cluster ? Method ? Measure Depending on the scale level, select the measure; convert variables with multiple categories into a set of binary variables and use matching coef? cients; standardize variables if necessary (on a range of 0 to 1 or A1 to 1). Two-step clustering: ? Analyze ? Classify ? Two-Step Cluster ? Distance Measure Use Euclidean distances when all variables are continuous; for mixed variables, use log-likelihood. ? Analyze ? Classify ?

Hierarchical Cluster ? Choose clustering algorithm Method ? Cluster Method (only hierarchical clustering) Use Ward’s method if equally sized clusters are expected and no outliers are present. Preferably use single linkage, also to detect outliers. Decide on the number of clusters Hierarchical clustering: Examine the dendrogram: ? Analyze ? Classify ? Hierarchical Cluster ? Plots ? Dendrogram (continued) 264 Table 9. 11 (continued) Theory 9 Cluster Analysis Action Draw a scree plot (e. g. , using Microsoft Excel) based on the coef? cients in the agglomeration schedule. Compute the VRC using the ANOVA procedure: ? Analyze ?

Compare Means ? One-Way ANOVA Move the cluster membership variable in the Factor box and the clustering variables in the Dependent List box. Compute VRC for each segment solution and compare values. k-means: Run a hierarchical cluster analysis and decide on the number of segments based on a dendrogram or scree plot; use this information to run k-means with k

clusters. Compute the VRC using the ANOVA procedure: ? Analyze ? Classify ? K-Means Cluster ? Options ? ANOVA table; Compute VRC for each segment solution and compare values. Two-step clustering: Specify the maximum number of clusters: ? Analyze ? Classify ? Two-Step Cluster ?

Number of Clusters Run separate analyses using AIC and, alternatively, BIC as clustering criterion: ? Analyze ? Classify ? Two-Step Cluster ? Clustering Criterion Examine the auto-clustering output. Re-run the analysis using different clustering procedures, algorithms or distance measures. Split the datasets into two halves and compute the clustering variables' centroids; compare ce