Faculty of Commerce - Papers (Archive)   Faculty of Business

2008

Market segmentation in tourism

Sara Dolnicar
University of Wollongong, s.dolnicar@uq.edu.au

Market segmentation in tourism

Abstract

Tourists are not all the same, they have different pictures of their ideal vacation. Tourists are heterogeneous. Market segmentation is the strategic tool to account for heterogeneity among tourists by grouping them into market segments which include members similar to each other and dissimilar to members of other segments. Both tourism researchers and tourism industry use market segmentation widely to study opportunities for competitive advantage in the marketplace. This chapter explains the foundations of market segmentation, discusses alternative ways in which market segments can be formed, guides the reader through two practical examples, highlights methodological difficulties and points to milestone publications and recently published applications of market segmentation in the field of tourism.

Chapter 16
Market Segmentation in Tourism

Sara Dolnicar, University of Wollongong, Australia

Acknowledgement

Synopsis

Tourists are not all the same, they have different pictures of their ideal
vacation. Tourists are heterogeneous. Market segmentation is the strategic tool to
account for heterogeneity among tourists by grouping them into market segments
which include members similar to each other and dissimilar to members of other
segments. Both tourism researchers and tourism industry use market segmentation
widely to study opportunities for competitive advantage in the marketplace.

This chapter explains the foundations of market segmentation, discusses
alternative ways in which market segments can be formed, guides the reader through
two practical examples, highlights methodological difficulties and points to milestone
publications and recently published applications of market segmentation in the field of
tourism.

Keywords: market segmentation; a priori; commonsense; a posteriori; data-driven;
post-hoc.

Market Segmentation

Every tourist is different. Every tourist feels attracted by different tourist destinations, likes to engage in different activities while on vacation, makes use of different entertainment facilities and complains about different aspects of their vacation. While all tourists are different, some are more similar to each other than others: many people enjoy culture tourism, many tourists like to ski during their winter holiday and many tourists require entertainment facilities for children at the destination. Acknowledging that every tourist is different and that tourism industry cannot possibly cater for each individual separately forms the basis of market segmentation.

Smith (1956) introduces the concept of market segmentation as a strategy. He states (p. 6) that "Market segmentation […] consists of viewing a heterogeneous market (one characterized by divergent demand) as a number of smaller homogeneous markets". When segmenting a market, groups of individuals are developed which are similar with respect to some personal characteristic. The particular personal characteristic with respect to which similarity is explored is the segmentation criterion or segmentation base. Segmentation criteria / bases can be socio-demographics (for instance, old versus young tourists), behavioral variables (skiers versus sightseers) or psychographic variables (tourists motivated by rest and relation versus those motivated by action and challenges).

Market segmentation can be applied by any unit operating in tourism industry: hotels, travel agencies, tourist attractions, restaurants, and local charities. In this chapter, a tourism destination is the entity for which market segmentation is conducted.

The benefit of market segmentation lies in a tourist destination being able to specialize on the needs of a particular group and become the best in catering for this group. In doing so the destination gains a competitive advantage because (1) competition can be reduced from the global market to tourism destinations specializing on the same segment (e.g., all ecotourism destinations), (2) efforts can be focused on improving the product in a specific way rather than trying to provide all things to all people at high cost (e.g., a family destination is unlikely to need extensive nightlife options), (3) marketing efforts can be focused by developing the most effective message for the segment targeted (e.g., a sun and fun message for young tourists traveling with friends) and by communicating the message through the most effective communication channel for the segment (e.g., in national geographic or other nature magazines for ecotourists), and finally, (4) tourist experiencing a vacation at a destination that suits their special needs are likely to be more satisfied with their stay and, consequently, revisit and advertise the destination among like-minded friends. Or, as Smith stated in his seminal paper (1956, p. 5): "market segmentation tends to produce depth of market position in the segments that are effectively defined and penetrated. The [organization that] employs market segmentation strived to secure one or more wedge-shaped pieced [of the market cake]."

The examples above demonstrate that the expected outcome from market segmentation is competitive advantage. Consequently, the aim of the actual segmentation task is to group tourists in the way that is of most managerial value. In order for a segment to be managerially useful a number of requirements should be fulfilled (for more detail on evaluation criteria of segments see Frank, Massy, and Wind 1972; Wedel and Kamakura 1998).

1. The segment should be distinct meaning that members of one segment should be as similar as possible to each other and as different as possible from other segments.

2. The segment should match the strengths of the tourism destination.

3. The segment should be identifiable. While female travelers can be identified very easily, identification of those visitors who are motivated by rest and relaxation may not be as simple.

4. The segment should be reachable in order to enable destination management to communicate effectively. For instance, surf tourists are likely to read surf magazines which could be used to advertise the destination.

5. A segment should be suitable in size. This does not necessarily imply that a bigger segment is better. A tourism destination may choose to target a small niche segment that represents a large enough market for the particular destination and has the advantage of having very distinct requirements.

The above criteria for the usefulness of segments have to be considered when one or more of many possible segments are chosen for active targeting.

Market segments can be derived in many different ways. All segmentation approaches can be classified as being either *a priori (commonsense)* segmentation approaches (Dolnicar 2004a ; Mazanec 2000) or *a posteriori (post hoc, data-driven)* segmentation approaches (Dolnicar 2004a; Mazanec 2000; Myers and Tauber 1977). The names are indicative of the nature of these two approaches. In the first case destination management is aware of the segmentation criterion that will produce a potentially useful grouping (commonsense) in advance, before the analysis is undertaken (a priori). In the second case destination management relies on the analysis of the data (data-driven) to gain insight into the market structure and decides after the analysis (a posteriori, post hoc) which segmentation base or grouping is the most suitable one.

COMMONSENSE SEGMENTATION

In the case of commonsense segmentation destination management informs the data analyst about the personal characteristics believed to be most relevant for splitting tourists into segments. The choice of personal characteristics can be driven by experience with the local market or practical considerations. Most tourism destinations, for instance, use country of origin as a segmentation criterion. They profile tourists from different countries of origin and develop customized marketing strategies for each country. Even if this method is not the most sophisticated, country of origin segmentation offers major practical advantages of taking such an approach: most countries of origins speak a different language which requires customized messages to be developed anyway, each country of origin has different media channels.

Commonsense segmentation has a long history in tourism research with many authors referring to it as profiling. As early as 1970 tourism researchers did investigate systematic differences between commonsense segments with a publication titled "Study Shows Older People Travel More and Go Farther" (author unknown) appearing in the Journal of Travel Research. A vast amount of commonsense segmentation studies have been published since and are continuing to be published.

Dolnicar (2004a) concludes that commonsense segmentation remains the most common form of segmentation study conducted in academic (and most likely also industry) tourism research: 53 percent of all segmentation studies published in the last 15 years in the main outlet for tourism segmentation research (the Journal of Travel Research) were commonsense segmentation studies. Recent examples include Kashyap and Bojanic (2000), who split respondents into business and leisure tourists and investigates differences in value, quality and price perceptions, Israeli (2002), who compares destination images of disabled and not disabled tourists, Klemm (2002), who profiles in detail one particular ethnic minority in the UK with respect to their vacation preferences, and McKercher (2002), who compares tourists who spend their main vacation at a destination with those who only stop on their way through. Other commonsense studies are discussed in Dolnicar (2005).

Typical examples of areas in which commonsense segmentation approaches are regularly used include profiling respondents based on their country of origin, profiling certain kinds of tourists (e.g., culture tourists, ecotourists) and profiling tourists who spend a large amount of money at the destination (big spenders). In fact, geographical segmentation such as grouping tourists by the country of origin were among the first segmentation schemes to be used (Haley 1968).

A step by step outline of commonsense segmentation is given in Figure 1. Commonsense segmentation consists of four distinct steps: first, a segmentation criterion has to be chosen. For example, destination management may want to attract tourists from Australia. Country of origin represents the segmentation criterion in this case. In Step 2 all Australian tourist become members of segment 1 and all other tourists (or a more specific subset of other countries of origin) become segment 2 members.

Step 1: **Selection** of the segmentation criterion
(e.g. age, gender, $ spent, country of origin)

Step 2: **Grouping** respondents into segments by assigning each respondent to the respective segment

---

Step 3: **Profiling** of segments by identifying in which personal characteristics segments differ significantly

---

Step 4: **Managerial assessment** of the usefulness of the market segments (and formulation of targeted marketing activities).
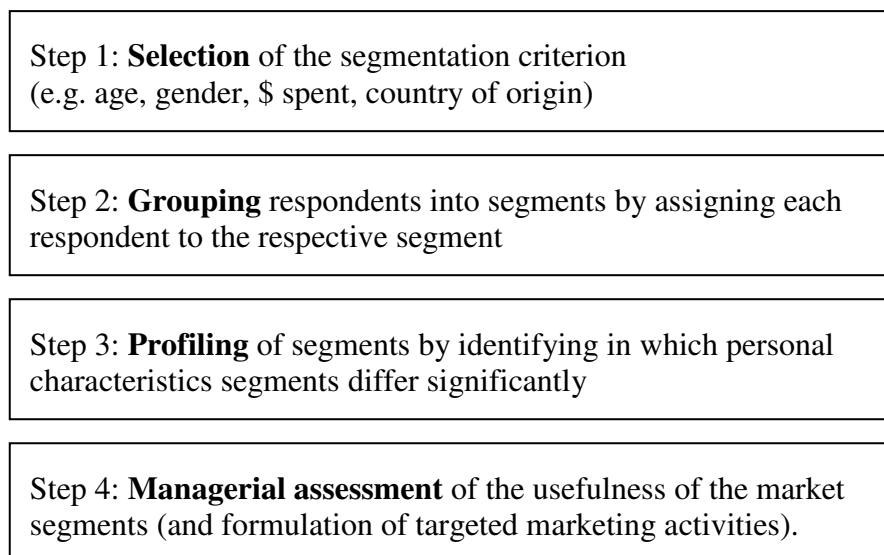

Figure 1: Steps in commonsense segmentation

Analyses of variance, t-tests, Chi-square tests or binary logistic regressions represent suitable techniques to test whether Australian tourists are significantly different from other tourists in Step 3. Note that the kind of test used depends on the number of characteristics that are tested and the scale of the variables. If many

characteristics are available in the data set the computation of independent tests for each characteristic overestimates the significance. Therefore, a Bonferroni correction is necessary on each p-value to account for this systematic overestimation, or researchers must choose methods, such as binary logistic regression, which automatically account for potential interaction effects between variables. The test chosen in Step 3 also needs to be appropriate for the scale of the data. If the profile regarding nominal (e.g., gender, type of vacation), binary (e.g., prior experience with the destination on a yes – no scale) or ordinal (e.g., income groups, level of expressed satisfaction) characteristics is tested, analysis of variance and t-tests are not the appropriate tests as they assume metric, normally distributed data. For some ordinal data this can be shown, but should be demonstrated before a test for metric data is applied.

Finally, in Step 4 destination management has to evaluate whether or not the commonsense segment of interest (e.g., Australian tourists) does represent an attractive market segment. This evaluation is made using the criteria outlined above. If the segment is attractive, destination management can proceed to customize the service to best suit the segment needs and develop targeted marketing activities which will enable most effective communication with the segment.

An Empirical Illustration: Country of Origin Segments

The National Guest Survey conducted in Austria in 1997 forms the basis of the empirical illustration. This type of Guest Survey data is very typical for the data sets available to destination management: a quota sample representative of the tourist population is taken and participating tourists are asked about their vacation preferences, travel behavior and personal characteristics.

Step 1: Criterion selection. For this illustration, country or origin is chosen as the segmentation criterion. Because Austria's tourism industry is highly dependent on German tourists, a profile of German tourists will be developed.

Step 2: Grouping. The total sample contains 6,604 respondents from 14 countries, of which 1,602 are from Germany. The variable containing the country of origin information is recoded to 0 and 1 where 1 indicates if the respondent comes from Germany and 0 indicates if the respondent comes from any of the other countries.

Step 3: Profiling. Analyses of variance and Chi-square tests are used commonly for profiling in tourism research. This approach is not wrong, but can lead to misinterpretations of the data in certain cases because independent tests are computed to assess the differences between the two segments for every single variable. Consequently, each of these tests does not account for interaction effects between all the variables that are tested and leads to an overestimation of significance. If a small number of variables is used for profiling this problem can be corrected through Bonferroni-correction of the p-values. The better solution is to use approaches where the differences between groups are compared for all variables simultaneously. This method does not limit the researcher with respect to the number of variables to be tested.

In this example, *binary* logistic regression is employed to profile the segment of German tourists because this statistical procedure allows for the dependent variable to only have two categorical options (membership in the segment of German tourists

versus membership in the other segment). The following variables were used to compare the two segments: socio-demographic personal characteristics (age, gender), behavioral characteristics (duration of stay, travel party, kind of holiday, accommodation, expenditures) and psychographic personal characteristics (travel motives).

Table 1 contains the regression coefficients resulting from the binary logistic regression. The Hosmer and Lemeshow test (0.248) indicates adequate fit of the model but the test statistics (Cox and Snell R-square = 0.101, Nagelkerke R-square = 0.143) and the percent of correctly predicted memberships (72.2 percent) for this particular illustration are not particularly good because only a very small number of variables was included which is insufficient to discriminate well between German tourists and tourists from other countries. Note that the aim is not an optimal prediction; instead, the objective is to determine statistically significant differences for the two segments. Significance values are provided in Table 1. At a significance level of 95% all p-values provided in the 6th column which are lower than 0.05 are significant.

Table 1: Binary logistic regression coefficients

	B	S.E.	Wald	df	p-value	Exp(B)
Age	0.02	0.00	25.56	1	**0.000**	**1.02**
Motive: Comfort (not important)	0.42	0.09	23.76	1	**0.000**	**1.53**
Motive: Culture (not important)	0.37	0.09	16.55	1	**0.000**	**1.45**
Motive: Fun (not important)	0.15	0.08	3.85	1	**0.050**	**1.17**
Motive: Nature (not important)	0.29	0.09	11.01	1	**0.001**	**1.33**
Motive: Beauty and Health (not important)	0.18	0.08	4.28	1	**0.039**	**1.19**
Motive: Relaxed atmosphere (not important)	-0.33	0.10	11.34	1	**0.001**	**0.72**
Motive: Cozy and familiar atmosphere (not important)	-0.51	0.09	34.13	1	**0.000**	**0.60**
Motive: Everything organized (not important)	0.21	0.10	4.38	1	**0.036**	**1.24**
Nights spent at destination	0.04	0.01	39.53	1	**0.000**	**1.04**
On vacation alone (not applicable)	0.75	0.22	11.76	1	**0.001**	**2.11**
On vacation with the family (not applicable)	0.36	0.17	4.52	1	**0.034**	**1.43**
On vacation with friends (not applicable)	0.41	0.10	18.82	1	**0.000**	**1.51**
Water and sun holiday (not applicable)	0.26	0.12	4.73	1	**0.030**	**1.30**
Hiking holiday (not applicable)	0.34	0.14	5.99	1	**0.014**	**1.40**
Culture holiday (not applicable)	0.31	0.08	14.57	1	**0.000**	**1.36**
Spa holiday (not applicable)	0.56	0.22	6.71	1	**0.010**	**1.75**
Sex (male)	0.34	0.07	20.19	1	**0.000**	**1.40**
Accommodation			83.63	6	**0.000**	
3 star hotel	1.33	0.18	54.40	1	**0.000**	**3.78**
1 or 2 star hotel	1.25	0.16	64.19	1	**0.000**	**3.48**
Holiday apartment	1.26	0.15	66.96	1	**0.000**	**3.53**
Private accommodation	1.21	0.16	55.65	1	**0.000**	**3.34**
Farmhouse accommodation	1.26	0.17	54.65	1	**0.000**	**3.54**
Camping ground	1.25	0.23	29.66	1	**0.000**	**3.48**
Expenditures per person per day	0.00	0.00	12.11	1	**0.001**	**1.00**
Constant	-6.12	0.44	189.42	1	**0.000**	**0.00**

Table 1 shows a number of significant differences exist between German and not German tourists spending their vacation in Austria. German tourists are older, comfort, culture, fun, nature, beauty and heath as well as a fully organized vacation

are more likely not to be important for German tourists, whereas a relaxed, cozy and familiar atmosphere matters to them (note that Exp(B) in the last column shows higher values for comfort, but given the coding of the variable as being not important the above interpretation results). German tourists spend more nights at the destination and are less likely to be traveling alone with family or friends. They are more likely to be male and to be found in lower category accommodation options.

Country of origin (COO) was chosen for this illustration because this method is one of the most frequently used commonsense segmentation approaches. However, a serious danger manifests with using country of origin as the segmentation criterion: if responses on rating scales (typically 5- or 7-point scales) form the basis for profiling, differences are very likely to be caused not only by differences in respondents attitudes, but also by differences in cross-cultural response styles. The data analyst must check for contamination of data by response styles before computing differences and possibly misinterpreting them. The problem of cross-cultural response styles can be avoided by using alternative answer formats. Pre-existing data can be checked for cross-cultural response styles. Practical recommendations are provided by Dolnicar and Grün (in press).

Step 1: Assessment. Although only a small number of variables are used for the illustration of commonsense segmentation, German tourists clearly demonstrate a distinct profile. German tourists are very easy to identify and are easy to reach. German tourist account for a substantial proportion of international recreational travel, the size of this market segment is therefore more than sufficient. In fact, an interesting subset to explore is German tourists, to develop a stronger competitive advantage. The final assessment criterion is match of the segment with the destination. This relationship cannot be assessed without a particular destination in mind, but given the profile information one can assume that a quiet relaxing destination which does not have much cultural or entertainment offers could match the needs of the segment of German tourists well.

DATA-DRIVEN SEGMENTATION

Data-driven segmentation studies do not have as long a history as commonsense segmentation studies do. Haley (1968) introduces data-driven market segmentation to the field of marketing. While acknowledging the value of geographic and socio-demographic information about consumers, Haley criticizes commonsense approaches as being merely descriptive rather than being based on the actual cause of difference between individuals and instead proposed to use information about benefits consumers seek to form market segments. This approach requires groups of consumers to be formed on the basis of more than one characteristic and, consequently requiring different statistical techniques to be used. As Haley (p. 32) states,"All of these methods relate the ratings of each respondent to those of every other respondent and then seek clusters of individuals with similar rating patterns."

About one decade after Haley has proposed data-driven market segmentation, tourism researchers adopted the method and published the first data-driven segmentation studies in tourism (Calantone, Schewe and Allen 1980; Goodrich 1980; Crask 1981; Mazanec 1984). A large number for data-driven segmentation studies has been published since with recent examples including work by Bieger and Lässer (2002), who construct data-driven segments among Swiss population on the basis of

travel motivations. This study represents data-driven segmentation in its pure form because no pre-selection of respondents takes place before the segmentation study is conducted. Contrarily Hsu and Lee (2002) use a subset of the tourist population as a starting point: only motor coach travelers. Among motor coach travelers they further segment tourists in a data-driven manner by exploring systematic differences in 55 motor coach selection attributes. Further examples are discussed in Dolnicar (2005).

The large number of data-driven segmentation studies published in the past two decades has led to a number of reviews of segmentation studies in tourism, some of which focus more on content, some on methodology.

Frochot and Morrison (2000) review benefit segmentation studies in tourism. They conclude that benefit segmentation leads to valuable insights in tourism research in the past, but recommend the following improvements: careful development of the benefit statements used as the segmentation base (some benefits are generic, but many are specific to the destination under study), informed choice of the timing (asking tourists before their vacation is less biased by the actual vacation experience), conduct benefit segmentation studies regularly to account for market dynamics and conduct them separately for different seasons.

Dolnicar (2002), based on a subset of studies reviewer by Baumann (2000), analyzes methodological aspects of data-driven segmentation studies in tourism concluding that only a small number of the available algorithms is used by tourism researchers who prefer either the hierarchical Ward's algorithm or the k-means partitioning algorithm. Dolnicar also identifies a number of problematic methodological standards that have developed in data-driven segmentation in tourism. To avoid data-driven segmentation studies that are of limited scientific and practical value it is important for data analysts and users to be aware of a number of basic principles upon which data-driven segmentation is based. These foundations are described in detail in the following section.

Foundations of data-driven market segmentation

Foundation 1: Market segmentation is an exploratory process. Many statistical techniques enable researchers to conduct test that provide one single correct answer for a research question. For instance, if an analysis of variance is conducted on destination brand image data, the test results inform the researcher whether or not there is a significant difference in the way respondents from different countries of origin perceive a destination. This test result is exactly the same, no matter how often the analysis is repeated. This method is not the case in data-driven market segmentation. Market segmentation is a process of discovery, an exploratory process. Aldenderfer and Blashfield (1984) refer to clustering, the algorithm typically used in data-driven market segmentation in tourism, as "little more than plausible algorithms that can be used to create clusters of cases." Each algorithm produces a different grouping and even repeated computations of one algorithm will not lead to the same segments. This point is very important to both researchers conducting data-driven market segmentation and managers using segmentation results. As a consequence, the choice of the segmentation algorithm and the parameters of the algorithm can and do have a major impact on the results. Data analysts must be aware of the fact that their selection of a data-driven segmentation procedure is "structure-imposing" (Aldenderfer and Blashfiled 1984) and that segmentation results from one algorithm

are unlike to have revealed the one and only true segmentation solution for any given data set.

Foundation 2: Market segments rarely occur naturally. The exploratory nature of market segmentation leads to a question which has rarely been discussed in marketing or tourism research: are market segments real and is the data analyst's aim to identify such naturally occurring segment or are market segments an artificial construction of groups for a particular purpose. Different authors take distinctly different positions on the matter. The seminal market structure analysis and market segmentation studies (Frank, Massy, and Wind 1972; Myers and Tauber 1977) imply that the aim of market segmentation is to find natural groupings. More recently, Mazanec (1997) and Wedel and Kamakura (1998) state explicitly that market segmentation typically means that artificial groupings of individuals are constructed.

Empirically both cases can occur and represent to extremes on the continuum of highly structured to not structured data sets. These two extreme options have been referred to as "true clustering" and "constructive clustering" by Dolnicar and Leisch (2001). Figure 2 illustrates the segmentation concepts available to data analysts in dependence of the structure of the empirical data being analyzed.
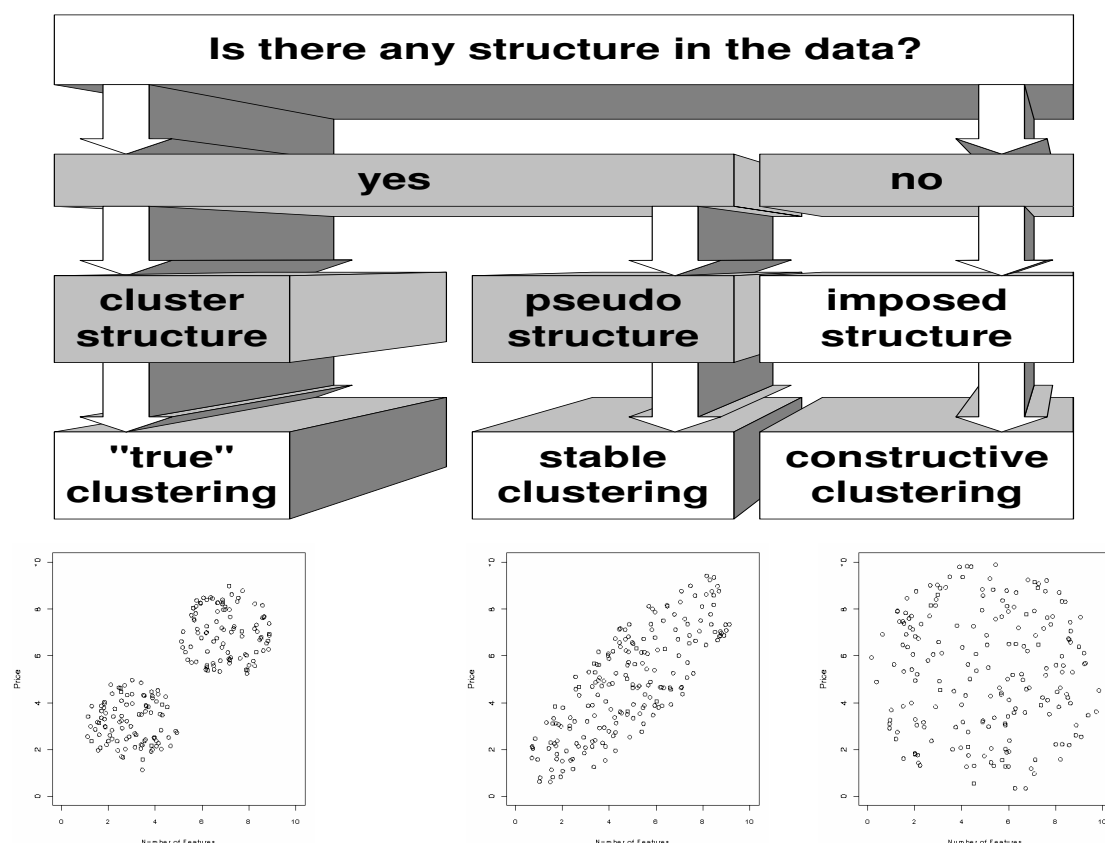

Figure 2: Alternative concepts of data-driven market segmentation (modified from Dolnicar & Leisch, 2001)

The bottom plots in Figure 2 represents a two dimensional empirical data set with two variables and could, for instance, be interpreted as the number of features tourist expect in a hotel room (e.g., TV, phone, and shoe cleaning kit; plotted on the x-

axis) and their willingness to pay a low or high price for the hotel. Optimally (illustrated by the bottom left hand plot) one segment would expect a large number of features and would be willing to pay a high price. The second segment would have low expectations about the number of features but would not be willing to pay a high room rate. In such a clear case of "true clustering", most algorithms and replications within the same algorithm would identify the correct groups. Managers could rely on the existence of these "true segments" and could target the big spenders by offering them a few more room features.

Unfortunately, the other extreme is more frequently encountered when human beings are the object of study: people have all levels of expectations regarding hotel room features and their willingness to pay ranges widely, including all levels of room rates (illustrated in the bottom right hand plot). Nevertheless, constructing a tourist segment of those who are willing to pay a high room rate is possible and managerially useful because marketing messages can be customized to these people and marketing expenses can be used more efficiently by advertising only through channels used by members of the "high room rate – high feature expectation" segment. In such a case every single segmentation computation would lead to a different grouping; no stability would be detectable across algorithms or repetitions. The implication is not that "constructive clustering" is wrong; instead, the inference is that managers must be informed about the nature of the segments they are targeting. If true segments are targeted, groupings of tourist should be used as identified in the segmentation study. If, however, constructed segments are used, they can be modified and improve to achieve the aim managers wish to achieve. For instance, managers may want to include another criterion, form a smaller or bigger group, split the "high room rate – high feature expectation" in two subgroup.

Finally, the last option illustrated in Figure 2 is referred to as "stable clustering" (bottom middle plot). In this case data are not entirely unstructured, but the structure does not represent distinct groups. The advantage of this data situation over the "constructive clustering" case is that – although segments are still artificially constructed – the data structure gives the analyst guidance with respect to the kind of artificial segments that should be constructed: if repeated computation with the same algorithm leads to similar results, the segmentation solution is more stable and should be preferred to an alternative segmentation solution which leads to large differences in the results emerging from repeated computations.

Foundation 3: One valid segment is enough. Most data-driven segmentation studies assess the validity of the final segmentation solution by profiling the resulting segments. If segments are distinct not only with respect to the segmentation base but also with respect to additional (external) criteria, segments can be seen to have external validity which increases their attractiveness for destination management. For instance, segments derived from travel motivations are more useful if they also differ in socio-demographic and media behavior, because they can be identified and communicated with more efficiently.

Typically, validity is assessed across all of the resulting segments, even if a tourist destination is only interested in one small niche market. If validity is low across all segments, a segmentation solution may be rejected although it may have contained one very distinct small segment which is highly externally valid. Please note that the entire segmentation solution is not necessarily of managerial interest. Validity assessment should reflect the managerial aim of the study.

Foundation 4: Market segmentation is not an independent strategic issue. Market segmentation is only one aspect of a tourist destination's strategy. This conclusion was already drawn by Smith (1956) when he first introduced market segmentation by stating that "Success in planning marketing activities requires precise utilization of both product differentiation and market segmentation as components of marketing strategy." Although the need to integrate segmentation into the total strategy of a tourist destination, most of the published data-driven segmentation studies in tourism treat segmentation as a separate issue and do not relate the results derived from segmentation to the positioning of the tourism destination or its competitive situation, although it is essential for successful implementation. Only if the destination can actually provide (positioning) for what the segment is seeking and this offering is distinctly different (differentiation) from competitors will market segmentation be more than an academic exercise and lead to a competitive advantage of a tourism destination. A method that automatically accounts for segmentation, positioning and competition is perceptual based market segmentation (PBMS, Mazanec and Strasser 2000; Buchta, Dolnicar, and Reutterer 2000). PBMS is based on tourists' evaluations of multiple destinations and preference information and enables the simultaneous analysis of segments, image positions and the extent of perceptual competition destinations face.

Foundation 5: Data-driven market segmentation is static, the market is dynamic. Any data-driven segmentation study can only provide insight into the market structure at the time of surveying tourists. Consequently, destination management needs to regularly repeat segmentation studies to identify chances in the segment structure of developments in specific segments that may be of particular value to the destination. A tracking framework for segmentation studies to be conducted over multiple periods of time has been proposed by Dolnicar (2004b).

To sum up, the implications of all the foundations outlined above are the following (1) data analysts and managers need to be aware of the exploratory nature of data-driven market segmentation and not over interpret the value of one single segmentation solution which was not based on thorough preliminary data structure analysis, (2) repeated computations of segmentation solutions can easily be undertaken to assess the stability of alternative solutions, (3) stability analysis will inform the data analyst and manager about the nature of the derived segmentation, whether it reveals true clusters, identified stable artificial groupings or represents an artificial construction of the most suitable grouping in a data set with very little structure, (4) it is not necessary for all segments derived from a segmentation solution to be valid. For a tourism destination searching for a niche, it is perfectly sufficient to have identified or constructed one market segment which has high external validity, (5) market segmentation is not independent. A successful market segmentation strategy is in line with the tourism destinations positioning and differentiation strategy, thus accounting for the particular strengths of the destination and the competitive environment, and (6) data-driven segmentation studies have to be repeated regularly to ensure validity of the insight gained into the market structure at any point in time.

Conducting data-driven market segmentation

A data-driven segmentation study contains all the components of a commonsense segmentation study. The way in which respondents are grouped is the

only difference between the commonsense and the data-driven approach: in commonsense segmentation one criterion is selected which usually is one single variable such as age or gender or high versus low levels of tourism spending. In data-driven segmentation a number of variables which ask respondents about different aspects of the same construct (e.g., a list of travel motives, a list of vacation activities) form the basis of segmentation and a procedure – in tourism research typically a clustering algorithm - is used to assign respondents to segments based on the similarity relationships between respondents. Figure 3 illustrates the additional steps needed for data-driven segmentation as steps 2a-2c.
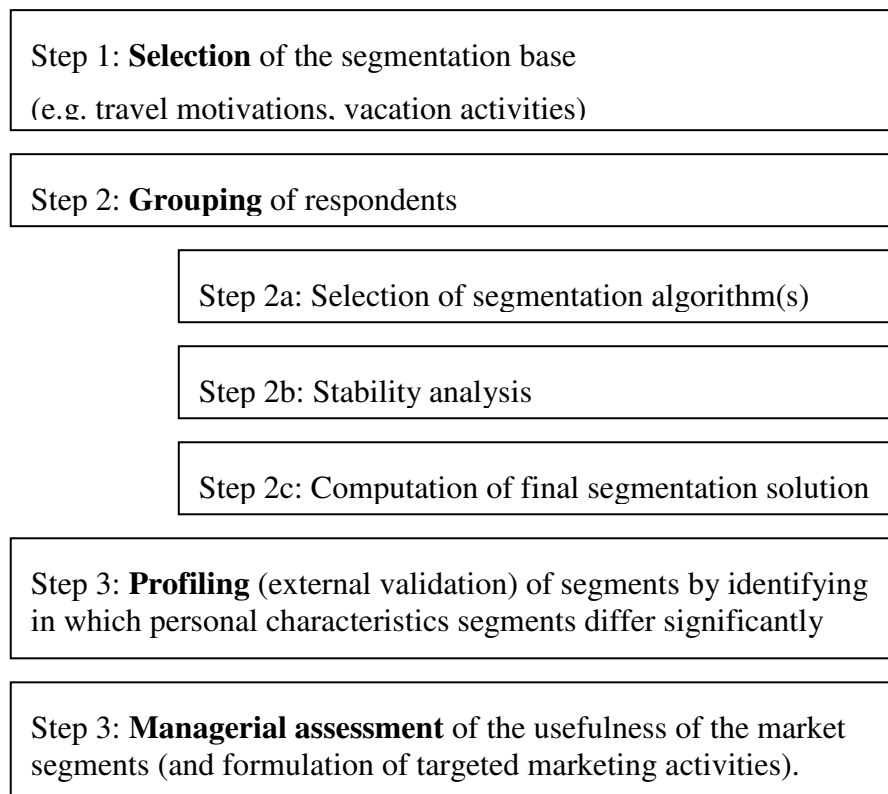
Step 1: **Selection** of the segmentation base
(e.g. travel motivations, vacation activities)

Step 2: **Grouping** of respondents

Step 2a: Selection of segmentation algorithm(s)

Step 2b: Stability analysis

Step 2c: Computation of final segmentation solution

Step 3: **Profiling** (external validation) of segments by identifying in which personal characteristics segments differ significantly

Step 3: **Managerial assessment** of the usefulness of the market segments (and formulation of targeted marketing activities).

Figure 3: Steps in data-driven segmentation

In step 2a the data analyst selects one or more segmentation algorithms. The predominant algorithms used in tourism research are k-means clustering and Ward's clustering. Ward's clustering is one form of hierarchical clustering procedures. Hierarchical – more precisely agglomerative hierarchical - clustering procedures determine the similarity between each pair of two respondents and then choose which two respondents are most similar and places them into a group. This process is repeated until all respondents are in one single group. The disadvantage of hierarchical algorithms is that they require computations of all pair-wise distances at each step which can be a limiting factor when working with very large data sets. The second most frequently used data-driven segmentation algorithm in tourism research is k-means clustering. K-means clustering is an algorithm from the family of partitioning techniques. This technique does not require the computation of all pair wise distances. Instead the number of segments to be derived has to be stated in advance. Random points drawn from the data set represent these segments. In each

step of the iterative procedure the distance between each of the respondents and the "segment representatives" is computed and the respondent is assigned to the segment that best represents his or her responses. For example, if a five segment solution is computed, only five distance computations have to be calculated using partitioning techniques as opposed to as many distance computations as there are respondents in the sample when using hierarchical techniques.

Although k-means and Ward's clustering dominate data-driven segmentation studies in tourism, a large number of other algorithms is available to the data analyst: a wide range of alternative clustering algorithms (Everitt, Landau, and Leese 2001), neural networks (e.g., Mazanec 1992; Dolnicar 2002), bagged clustering (e.g., Dolnicar and Leisch 2003), latent class analysis (e.g., Van der Ark and Richards 2006), and finite mixture models (Wedel and Kamakura 1998).

When selecting an algorithm the data analyst should be aware of the advantages and disadvantages of the alternative methods and in particular the way in which they are known to impose structure on data. Most clustering algorithms allow the data analyst to define which distance measure should be used. Again, a large number of alternative distance measures are available. The data analyst has the responsibility to select a distance measure suitable for the data scale. For instance, metric and binary data can be analyzed using Euclidean distance. This choice is not necessarily the case for ordinal data. For a detailed discussion of alternative distance measures see Everitt, Landau, and Leese (2001).

Another point that should be noted while discussing the selection of a suitable clustering algorithm is the term "factor-cluster segmentation" which appears to have developed in tourism research. Researchers using this approach typically select a large number of items, conduct factor analysis to reduce a large number of items to a smaller number of factors and subsequently use factor scores as the basis for segmentation. This approach has two effects: (1) the original items are actually not used to segment. Consequently, resulting segments cannot be interpreted using the original items, because they emerged from a heavily transformed data space. (3) Factor analyses typically explain between 50 and 60 percent of the information contained in the original items. Conducting factor analysis before clustering essentially means that 40 to 50 percent of information is lost. Direct clustering of original items is therefore preferable if the aim of the segmentation study is to develop segments based on the questions asked in the survey (benefits, motivations, and behavior). Sheppard (1996) compares cluster analysis with factor-cluster analysis methods and concludes that factor-cluster analysis is not suitable if the study's aim is to examine heterogeneity among tourists; factor analysis may be a valuable approach for the development of instruments for the entire population assuming homogeneity. Arabie and Hubert (1994) are less diplomatic by stating that "`tandem´ clustering is an outmoded and statistically insupportable practice" because the nature of the data is changed dramatically through a factor analytic transformation before segments are explored.

Data analysts also should keep in mind that the number of variables that can be analyzed with a sample of a certain size is limited. Although there are no specific rules for non-parametric procedures, a rule of thumb proposed by Formann (1984) provides some helpful guidance: for the case of binary data (yes no questions) the minimal sample size should include no less than 2^k cases (k = number of variables), preferably $5*2^k$ of respondents.

Finally, the most unresolved question in market segmentation remains how to select the number of segments that best represents the data or most suitably splits respondents into managerially useful segments. A large number of heuristics exist to assess the optimal number of clusters but comparative studies show that no single one of these indices is superior to the others. If the data is well structured, the correct number of clusters will be identified by most heuristic procedures. If the data is not well structured, which is typically the case in the social sciences, heuristics are not helpful to the data analyst. The approach the author finds most useful is based on the above mentioned concepts of segmentation (Figure 2) where data structure is the driving force and stability is the criterion. To determine the number of clusters using the stability criterion, a number of repeated computations are conducted and the agreement across alternative solutions is assessed. The number of clusters that leads to the most stable results over repeated computations wins.

An empirical illustration: behavioral segments

Once again the Austrian Guest Survey serves as the underlying empirical data set for the illustration. This illustration is limited to the additional Steps 2a, 2b and 2c as outlined in Figure 3. The following ten vacation activities form the basis of the behavioral segmentation analysis: cycling, swimming, going to a spa, surfing / sailing, boat riding, relaxing, going out in the evenings, shopping, sightseeing and visiting theatres, musicals and operas. The sample contains 6,604 respondents. Based on the recommendation by Formann (1984), the minimum sample size for ten variables is 1,024, optimally 5,120 respondents would be available. Our data set complies with both recommendations.

Step 2a: Selection of segmentation algorithm(s). The algorithm used is a topology representing network (TRN, Martinetz, and Schulten 1994). Topology representing networks are neural networks similar to self organizing feature maps (Kohonen 1984). The algorithm is similar to the k-means algorithm but allows for neighboring cluster centroids to be updated as well thus leading to a topological representation of resulting segments in space. The reason for this choice is that the data set is large and partitioning algorithms consequently represent the computationally more efficient option. Furthermore, results from Monte Carlo studies with artificial data sets have shown that TRNs outperform a number of other partitioning algorithms, including k-means in identifying the correct structure of the data (Buchta et al. 1997). The software to compute TRNs is available at http://tourism.wu-wien.ac.at/cgi-bin/ift.pl?charly/http/software/contents.html. Please note, however, that a wide range of algorithms is available. All algorithms lead to the assignment of each respondent to one segment.

Euclidean distance computes the differences between cluster centroids (segment representatives) and respondents. This analysis is permissible because the data set is binary. Before the computation 100 random draws of starting points are undertaken. These randomly drawn staring points become the initial cluster centroids. The best of 1,000 is used to start the training process which is undertaken for 100 iterations.

Step 2b: Stability analysis. Fifty replications are computed for clusters numbers from three to ten. Two and three clusters are not included in this stability comparison because they are typically dominated by two clusters: those who say yes

more frequently and those who say no more frequently, thus not providing any interesting insights about the resulting clusters to management. The stability of each number of clusters over the 50 replications is assessed by computing how many pairs of respondents are assigned to the same segment. The "percentage uncertainty reduction" value in the TRN software is an indicator of this value. The higher the percentage the more stable are the results emerging from the repeated computations. Table 3 contains the uncertainty reduction values for all computations.

Table 2: Stability of solutions ranging from 4 to nine segments

Number of clusters	Number of repeated calculations	Percent uncertainty reduction	Improvement in percent uncertainly reduction
4	50	73.79	
5	50	78.04	4.25
6	**50**	**84.82**	**6.78**
7	50	86.89	2.07
8	50	87.83	0.94
9	50	89.69	1.86

The stability is increasing as higher numbers of clusters are developed. The highest improvement in stability occurs from the five to the six segment solution. Consequently, the 6 segment is chosen for interpretation. Stability values do not, however, indicate that the six segment solution is the far superior solution, indicating that the data-driven segmentation undertaken is likely to be of "stable clustering" or even "constructive clustering" nature.

Step 2c: Computation of final segmentation solution. The basis of interpreting data-driven market segments is the comparison of segment averages and sample averages for all items in the segmentation base. The closer the segment average to the sample average, the less distinct is the segment. Figure 4 provides profile charts for all six resulting market segments. The black columns indicate the segment average for each of the items in the segmentation base, the dark grey horizontal bars depict the sample average.
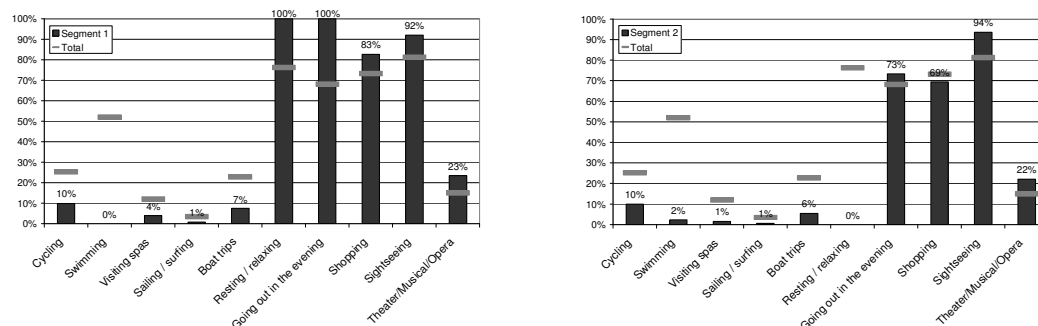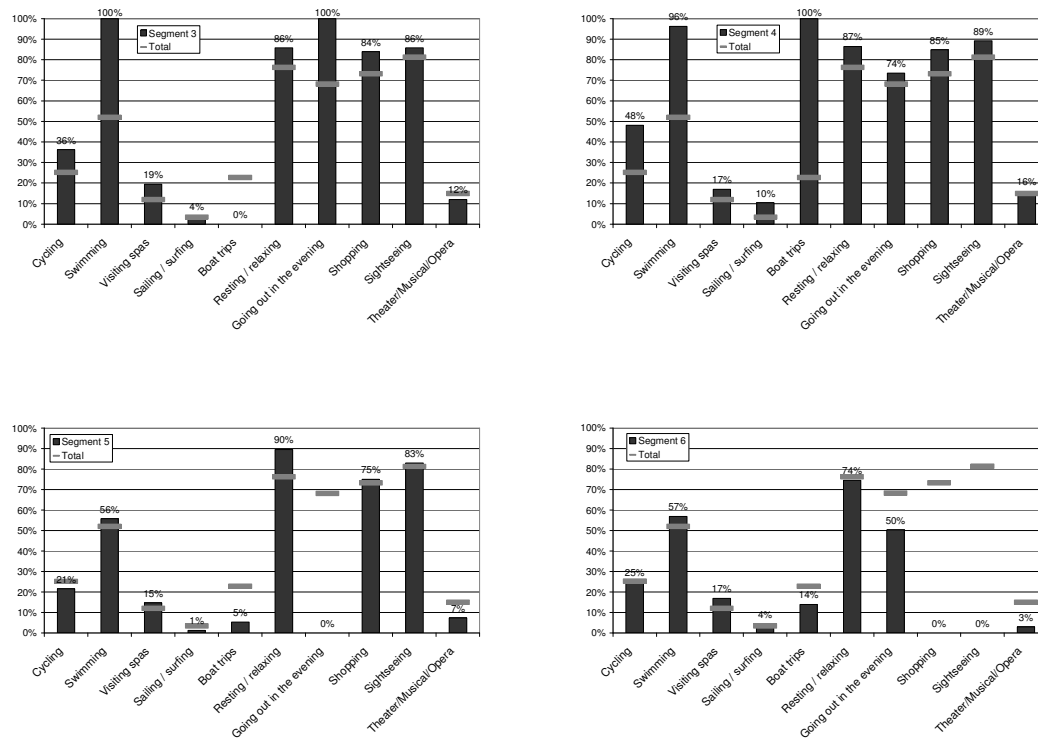
Figure 4: Data-driven segment profiles

As can be seen the market segments resulting from the six segment solution are quite distinct. Members of Segment 1 (represents 21 percent of the sample) are interested in resting and relaxing, but they also engage in cultural activities and shopping. They practically do not engage in typical summer vacation activities, such as swimming, cycling or going on boat trips. Segment 2 members (15 percent) are interested in cultural activities, shopping and going out in the evening. Resting is not part of their vacations; their activity profile indicates that they are highly active when on vacation. Segment 3 (19 percent) represents tourists with a more traditional pattern of summer activities. They engage in cycling and swimming more frequently than the total sample does, but culture, going out in the evenings and shopping are part of the vacation program as well, as is resting and relaxing. Segment 4 (18 percent) members state to engage in all vacation activities more frequently than the sample average. This segment should be interpreted with case as it could be reflecting a so-called acquiescence (yes saying) response style. Segment 5 (19 percent) is really most interested in resting and relaxing. Segment 6 (8 percent) engages in summer sports activities, but is not interested in cultural activities at all.

Some researchers have in the past conducted tests (such as analyses of variance or Chi-squared tests) to assess whether the differences between segments along items of the segmentation base are significant and have argued that his approach can be used to validate the segmentation solution. Because segmentation algorithms construct the most distinct solution, this method is not an appropriate test of validity. An extremely surprising outcome would be no significant differences between resulting segments. Validity of solutions is checked using additional variables that were not exposed to the segmentation algorithm. This checking is done in Step 3 of

the data-driven segmentation process. Steps 3 and 4 are not illustrated for the data-driven segmentation as the same approach is used as explained in detail for the case of commonsense segmentation.

OTHER APPROACHES TO CREATING MARKET SEGMENTS

Although the majority of market segmentation studies in tourism are typically classified as being commonsense segmentation studies or data-driven segmentation studies, combinations of both approaches are possible and may represent a useful alternative for tourism managers to explore potentially attractive target segment for their purposes. Dolnicar (2004a) gives an overview of such alternative segmentation approaches. The classification of these approaches (left side of Figure 5) assumes that a two-stage process is taken where the data analyst first creates a commonsense or a data-driven segmentation and then continues with an additional analysis afterwards. For instance, destination management could first split tourists based on their country of origin and then in the second step either (1) search for distinct groups differing in their travel motivations (which would represent a Concept 5 segmentation) or (2) split respondents into first time and repeat visitors (Concept 3).
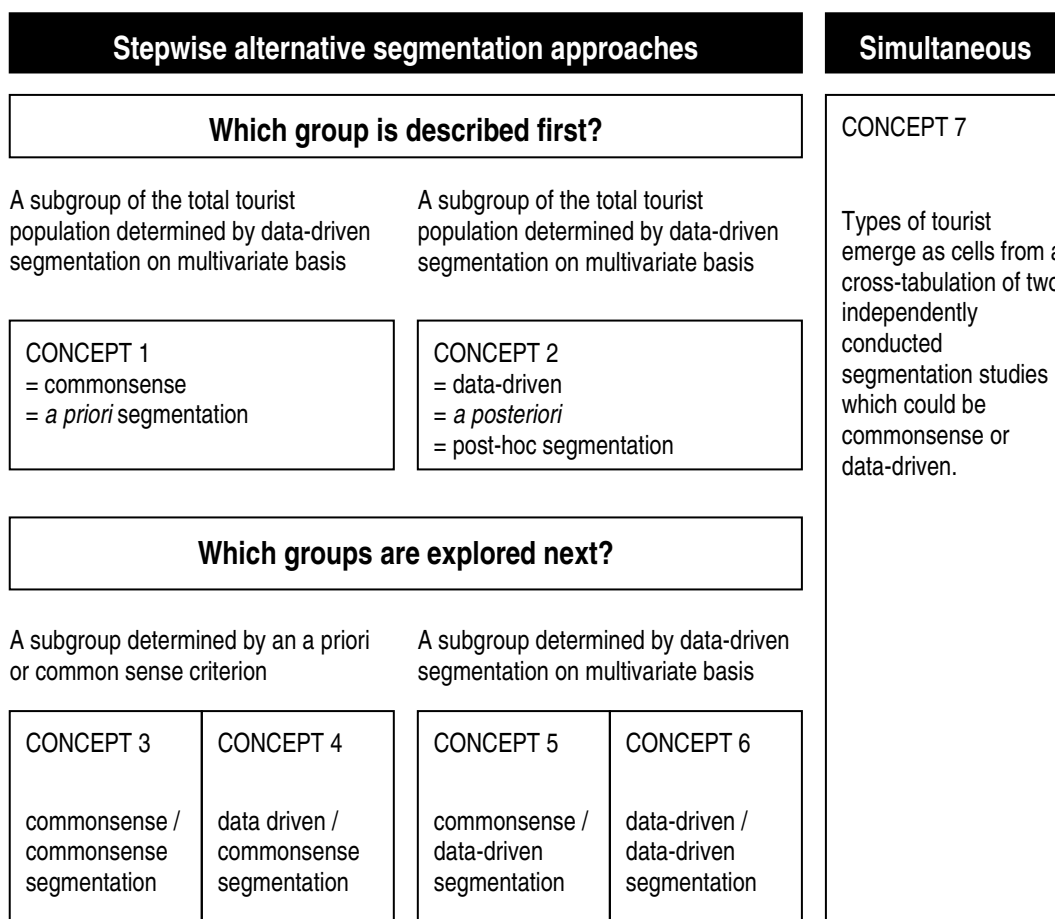
Stepwise alternative segmentation approaches		Simultaneous
Which group is described first?		CONCEPT 7
A subgroup of the total tourist population determined by data-driven segmentation on multivariate basis	A subgroup of the total tourist population determined by data-driven segmentation on multivariate basis	Types of tourist emerge as cells from a cross-tabulation of two independently conducted segmentation studies which could be commonsense or data-driven.
CONCEPT 1 = commonsense = *a priori* segmentation	CONCEPT 2 = data-driven = *a posteriori* = post-hoc segmentation	
Which groups are explored next?		
A subgroup determined by an a priori or common sense criterion	A subgroup determined by data-driven segmentation on multivariate basis	
CONCEPT 3 commonsense / commonsense segmentation	CONCEPT 4 data driven / commonsense segmentation	CONCEPT 5 commonsense / data-driven segmentation	CONCEPT 6 data-driven / data-driven segmentation	

Figure 5: A systematics of market segmentation approaches (modified from Dolnicar, 2004a)

Of course, managers may be interested in exploring combinations of simultaneously constructed market segments. Combination methods are done by conducting two independent segmentation studies based on different segmentation bases and then simply cross-tabulating the resulting groups. For instance, destination management could construct segments based on motives and segments based on vacation activities independently based on the same data set and then investigate whether these two segmentations are associated and result in interesting vacation types. One example for such a simultaneous segmentation study is provided by Dolnicar and Mazanec (2000).

Note that while such alternative segmentation approaches are useful in exploring potentially interesting target segments they can also be used to externally validate segments. For instance, if country of origin is used as an a priori segmentation criterion, researchers could investigate whether segments of tourists who differ with respect to their tourism motivations are associated with the country of origin grouping.

CONCLUSION

Market segmentation is a strategy any entity in the tourism industry can use to strengthen their competitive advantage by selecting the most suitable subgroup of tourists to specialize on and target.

A wide variety of alternative techniques can be used to identify or construct segments. Approaches range from simple commonsense segmentations (where tourists are split on the basis of a predefined personal characteristic) to multidimensional data-driven approaches where a set of tourist characteristics is used as the basis for grouping. Once tourists are grouped using the correct and most suitable analytical techniques the resulting segmentation solution has to be assessed by the users (tourism managers) who will not only evaluate the segmentation solution per se but also the fit of potentially interesting segments with the strengths of the tourism destination.

Tourism managers can benefit from market segmentation by using it actively as a method of market structure analysis. In doing so, they can gain valuable insight into the market and specific sections of the market and identify the most promising strategy to gain competitive advantage. Typically such a strategy will not only require market segmentation, but also product positioning. Both approaches will have to be evaluated in view of competitors' segmentation and positioning choices to be successful. Segmentation solutions should be computed regularly to ensure that current market structure is captured.

REFERENCES

Aldenderfer, M.S., & Blashfield, R.K. (1984). Cluster Analysis. Beverly Hills: Sage Publications.

Arabie, P., & Hubert, L. (1994). Cluster Analysis in Marketing Research. In R. Bagozzi (Ed.) Advanced methods of marketing research (160-189). Cambridge: Blackwell.

Baumann, R. (2000). Marktsegmentierung in den Sozial- und Wirtschaftswissenschaften: eine Metaanalyse der Zielsetzungen und Zugänge. Diploma thesis at Vienna University of Economics and Management Science. Vienna.

Bieger, T., & Lässer, C. (2002). Market Segmentation by Motivation: The Case of Switzerland. Journal of Travel Research, 41 (1), 68-76.

Buchta, C., Dimitriadou, E., Dolnicar, S., Leisch, F., & Weingessel, A. (1997). A Comparison of Several Cluster Algorithms on Artificial Binary Data Scenarios from Travel Market Segmentation. Working Paper # 7, SFB 'Adaptive Information Systems and Modelling in Economics and Management Science', Vienna.

Buchta, C., Dolnicar, S., & Reutterer, T. (2000). A Nonparametric Approach to Perceptions-Based Market Segmentation: Applications. Series: Interdisciplinary Studies in Economics and Management, Volume II. Springer, Berlin.

Calantone, R., Schewe, C., & Allen, C.T. (1980). Targeting Specific Advertising Messages at Tourist Segments. In D.E. Hawkins, E.L. Shafer, and J.M (Eds.), Tourism Marketing and Management (pp. 133-147). Washington D.C: George Washington University

Crask, M. (1981). Segmenting the Vacationer Market: Identifying the Vacation Preferences, Demographics, and Magazine Readership of Each Group. Journal of Travel Research, 20, 20-34.

Dolnicar, S., & Grün, B. (in press) Cross-Cultural Comparisons - Assessing analytical robustness. International Journal of Tourism, Culture, and Hospitality Research.

Dolnicar, S., & Leisch, F. (2001). Knowing What You Get - a Conceptual Clustering Framework for Increased Transparency of Market Segmentation Studies. Marketing Science, Wiesbaden, Germany, 5. – 8. July.

Dolnicar, S. & Leisch, F. (2003). Winter Tourist Segments in Austria – Identifying Stable Vacation Styles for Target Marketing Action. Journal of Travel Research, 41 (3), 281-193.

Dolnicar, S. (2002). Activity-Based Market Sub-Segmentation of Cultural Tourists. Journal of Hospitality and Tourism Management, 9 (2), 94-105.

Dolnicar, S. (2004a). Beyond "Commonsense Segmentation" – a Systematics of Segmentation Approaches in Tourism. Journal of Travel Research, 42 (3), 244-250.

Dolnicar, S. (2004b). Tracking Data-Driven Market Segments. Tourism Analysis, 8 (2-4), 227-232.

Dolnicar, S. (2005). Empirical Market Segmentation: What You See Is What You Get. In: Theobald, W. (Ed.), Global Tourism, The Next Decade, 3rd ed (pp. 309-325). Oxford: Butterworth-Heinemann.

Dolnicar, S., & Mazanec, J. (2000). Vacation Styles and Tourist Types: Emerging New Concepts and Methodology. In: Gartner, W.C. (Ed.), Trends in Outdoor Recreation and Tourism (pp 245-255). New York: CAB International.

Everitt, B.S., Landau, S., & Leese, M. (2001). Cluster Analysis. London: Arnold.

Formann, A.K. (1984). Die Latent-Class-Analyse: Einführung in die Theorie und Anwendung. Weinheim: Beltz.

Frank, RE., Massy, WF., & Wind, Y (1972). Market Segmentation, Englewood Cliffs: Prentice-Hall.

Frochot, I., & Morrison, A. M. (2000). Benefit segmentation: A review of its application to travel and tourism research. Journal of Travel and Tourism Marketing, 9 (4), 21-45.

Goodrich, J. (1980). Benefit Segmentation of US International Travelers: An Empirical Study with American Express. In D.E. Hawkins., E.L. Shafer., and J.M., (Eds.), Tourism Marketing and Management (pp. 133-147). Washington D.C: George Washington University.

Haley, R.J. (1968). Benefit Segmentation: A Decision-Oriented Research Tool. Journal of Marketing, 32: 30-35.

Hsu, C., & Lee, E. (2002). Segmentation of Senior Motorcoach Travelers. Journal of Travel Research, 40 (4), 364-374.

Israeli, A. (2002). A Preliminary Investigation of the Importance of Site Accessibility Factors for Disabled Tourists. Journal of Travel Research, 41 (1), 101-104.

Kashyap, R., & Bojanic, D.C. (2000). A Structural Analysis of Value, Quality, and Price Perceptions of Business and Leisure Travelers. Journal of Travel Research, 39 (1), 45-51.

Klemm, M.S. (2002). Tourism and Ethnic Minorities in Bradford: The Invisible Segment. Journal of Travel Research, 41 (1), 85-91.

Kohonen, T. (1984, 3rd ed. 1988). Self-Organization and Associative Memory. New York: Springer.

Martinetz, Th., & Schulten, K. (1994). Topology Representing Networks. Neural Networks 7 (5), 507-522.

Mazanec, J. (Ed.) (1997) International City Tourism: Analysis and Strategy. London: Pinter/Cassell.

Mazanec, J.A. (1984). How to detect Travel Market Segments: A Clustering Approach. Journal of Travel Research, 23 (1): 17-21.

Mazanec, J.A. (1992). Classifying Tourists into Market Segments: A Neural Network Approach. Journal of Travel & Tourism Marketing, 1, 39-59.

Mazanec, J.A. (2000). Market Segmentation. In: J. Jafari (Ed), Encyclopedia of Tourism. London: Routledge.

Mazanec, J.A., & Strasser, H. (2000). A Nonparametric Approach to Perceptions-Based Market Segmentation: Foundations. Berlin: Springer.

McKercher, B. (2001). A Comparison of Main-Destination Visitors and through Travelers at a Duel-Purpose Destination. Journal of Travel Research, 39 (4), 433-41.

Myers, J.H., & E. Tauber (1977). Market structure analysis. Chicago: American Marketing Association.

Sheppard, A.G. (1996). The sequence of factor analysis and cluster analysis: Differences in segmentation and dimensionality through the use of raw and factor scores. Tourism Analysis, 1, 49-57.

Smith, W. (1956). Product Differentiation and Market Segmentation as Alternative Marketing Strategies. Journal of Marketing, 21, 3-8.

Van der Ark, L. A., & Richards, G. (2006). Attractiveness of cultural activities in European cities: A latent class approach. Tourism Management, 27 (6), 1408-1413

Wedel, M., & Kamakura, W. (1998). Market Segmentation - Conceptual and Methodological Foundations. Boston: Kluwer Academic Publishers.