Tourism segmentation by consumer-based variables

S. Dolnicar* B. Kemp†

*University of Wollongong, sarad@uow.edu.au

†University of Wollongong, byron_kemp@uow.edu.au

This book chapter was originally published as Dolnicar, S & Kemp, B, Tourism Segmentation by Consumer-Based Variables, in Kozak, M & Decrop, A (eds) Handbook of Tourist Behavior: Theory & Practice, Routledge, New York, 2009, 177-194. Original book information available here

Tourism segmentation by consumer-based variables

Sara Dolnicar and Byron Kemp

Marketing Research Innovation Centre (MRIC)
School of Management and Marketing
University of Wollongong
Northfields Ave
2522 Wollongong NSW Australia

sara_dolnicar@uow.edu.au

byron_kemp@uow.edu.au

INTRODUCTION

The basis for successful marketing is to understand and satisfy consumer needs. Sometimes it is even possible to satisfy one individual customer's needs. In the tourism industry an individually customized tourism experience can be developed, but the market for such high-end tourism products is small. This does not, however, mean that the only alternative is to appeal to the mass market. The intermediate solution is to understand which groups of tourists have similar needs and develop tourism products that match group needs. This approach is referred to as market segmentation.

The aim of this chapter is to analyze market segmentation studies in tourism research over the past decade, review recent prototypical examples of different segmentation approaches and discuss theoretical and methodological issues related to market segmentation studies. Recommendations are presented that provide guidance to researchers and students with respect to how to best avoid potential pitfalls that may lead to misinterpretations of segmentation solutions and, consequently, sub-optimal strategic decisions.

SEGMENTING TOURISTS BY CONSUMER-BASED VARIABLES

Market segments are the result of splitting individuals according to a pre-defined rule. Consequently, many possible segmentation solutions exist for any given problem, depending on the pre-defined rule selected. Different segmentation solutions result from (1) different segmentation bases and (2) different segmentation methods. Figure 1 provides an overview of possible segmentation methods, including both conventional and alternative approaches.

Market segments can be defined using different segmentation bases (Wedel and Kamakura, 1998): tourists can be split into groups based on their country of origin, which arguably represents the most common market approach in tourism. Socio-demographic variables are frequently used in tourism as well: destinations may specialize in family vacations, thus trying to attract people with specific socio-demographic characteristics.

While geographic and socio-demographic segmentation is very popular in tourism, behavioral and psychographic segmentation criteria have received increasing attention since Haley (1968) introduced the concept of benefit segmentation. Behavioral variables of interest include ways in which different tourists organize their vacation (e.g. travel agent versus online) or tourists' vacation activities, which have direct impact on the tourism product design. Psychographic bases include benefits sought, travel motivations and destination preferences. All these segmentation bases can be referred to as consumer-based variables because they are characteristics of each individual consumer: having or not having a specific trait leads to the classification of consumers into different market segments.

The approaches most frequently stated to be used in tourism research are referred to as Concept 1 and Concept 2 in Figure 1. Concept 1 is the *a priori* (Mazanec, 2000) or *commonsense* (Dolnicar, 2004a) approach where one single splitting criterion is selected in advance and consumers are split based on their characteristic of the respective consumer-based variables such as young versus old tourists (e.g. Reece, 2004), female versus male travelers (e.g. Kim et al., 2007) or countries of origin (e.g. McCleary et al., 2006). The advantage of this approach is that it is methodologically simple; possible disadvantages are that management may select an unsuitable or sub-optimal splitting criterion, and that the splitting criterion may be used by many destinations or businesses, thus not providing a good basis for distinct image building and competitive advantage.

Concept 2 in Figure 1 is referred to as *post-hoc* (Myers & Tauber, 1977), *a posteriori* (Mazanec, 2000) or *data-driven* segmentation (Dolnicar, 2004a). The segmentation base is a set of variables. The advantage of this approach is that consumer-based variables, hypothesized to be associated more closely to destination or tourist business choice (such as benefits sought), can be used to determine market segments. Such segments may enable a more unique differentiation and thus offer a competitive advantage. The disadvantage is that the identification (Frank, Massy & Wind, 1972; Myers & Tauber, 1977) or construction (Mazanec, 1997; Wedel & Kamakura, 1998) of segments is achieved by using statistical techniques involving methodological parameters to be chosen by the researcher which have a major impact on the final segmentation solution. These decisions will be discussed in detail below.

Alternative segmentation approaches are rarely declared as such. They involve – either implicitly or explicitly - two grouping steps. Concept 3 refers to the situation where both groupings are made based on a single criterion (e.g. senior travelers among cultural tourists are profiled). In Concept 4 studies, the first grouping is data-driven, followed by a commonsense segmentation (e.g. repeat visitors among vacation activity segments). Concept 5 studies start with a commonsense segmentation and then proceed to construct a data-driven grouping (e.g. benefit segments among German tourists). Finally, Concept 5 studies represent a sequence of two data-driven segmentation studies (e.g. vacation activity segments further segmented by benefits sought).

Note that studies initially appearing to be Concept 2 often turn out to be Concept 5 studies because the starting point is a subset of the total population. For instance, Sung (2004) used a subset of the population (the adventure traveler market) as the basis for a data-driven segmentation based on variables including socio-demographics; psychographics and behavioral elements.

When Concept 3, 4, 5 and 6 studies are conducted simultaneously rather then subsequently they are referred to as Concept 7 studies. For example, an activity-based and a benefits-based segmentation are computed independently and then cross-tabulated to investigate resulting vacation styles (e.g. Dolnicar & Mazanec, 2000).

Finally, response-based approaches (Concept 8) have received significant attention in the marketing literature (Wedel & Kamakura, 1998), but have not been widely adopted in tourism. The difference between segmentation based on consumer characteristics and response-based segmentations is that the latter use consumer responses to specific marketing stimuli as the starting point. Thus, response based segmentation is determined not only by consumer characteristics, but by the interaction between marketing stimuli and consumer characteristics.

[INSERT FIGURE 1]

MARKET SEGMENTATION IN TOURISM

Due to the fact that a tourist population can be split in an endless number of ways, any of the segmentation approaches outlined in Figure 1 could suffer from the selection of a sub-optimal segmentation base. The selection of the segmentation base is a step in the segmentation process which has not received much research attention. One way to assess alternative segmentation bases would be to sun simulations with alternative bases and, in so doing empirically determine which segmentation base to choose. This, however, is not typically done. Usually the segmentation base is chosen in advance, based on managerial experience.

The second danger associated with all segmentation approaches at the profiling stage of the process is that group differences are frequently tested using independent tests for all variables of interest. Typically this is done by computing analyses of variance to test whether segments differ in metric variables (such as age, income, expenditure) and by computing Chi-squared tests to assess whether differences between market segments in nominal, binary or ordinal variables (such as gender, activities engaged in on vacation, country of origin) are significant. Because the same data set underlies all these tests, possible interactions between the variables are ignored, thus overestimating the significance of differences. This can lead to serious misinterpretations: marketing managers could falsely assume that the segments are highly distinct from each other in many ways. Correction for multiple testing or the use of a test that automatically accounts for multiple variables being tested simultaneously may, however, reveal that segments do not vary much at all and possible do not even represent suitable target groups.

A third general problem is that changes over time are rarely accounted for. A segmentation solution is developed at a certain point in time and may then be used as the basis for strategic marketing for many years without reviewing changes in the segments over time. Approaches to take changes over time into account have been proposed by Wedel and Kamakura (1998) in the general area of market segmentation and Dolnicar (2004b) in the tourism context.

A final issue is the need to integrate segmentation strategies into the overall marketing strategy (including product positioning, segmentation and competition). Only when all three strategic tools are well aligned, can the optimal results be achieved . Optimally, a destination or a tourism business conveys an image which is distinctly different from the image of competitors and is in line with the market segment targeted. Typically, however, segmentation studies focus exclusively on identifying the most attractive segment/s. Rarely are positioning and competition issues taken into consideration. One exception is a segmentation method which uses brand image of tourists as the segmentation base and in doing so, analyses all three aspects simultaneously. To date this method, referred to as Perceptions Based Market Segmentation (Mazanec & Strasser, 2000; Buchta, Dolnicar & Reutterer, 2000) or Perception-Based Analysis (Mazanec & Strasser, 2007) has not been widely adopted.

The following three sections review the predominant forms of segmentation studies by consumer-based variables in tourism. For this purpose, segmentation studies conducted in the past 10 years that were published in the Annals of Tourism Research, Tourism Management, Journal of Travel Research, Journal of Sustainable Tourism, and Journal of Travel and Tourism Marketing were reviewed. Each study was coded as one case in SPSS including the variables sample size, type of segmentation, type of segmentation base, number of variables, data format, data structure investigation, data pre-processing, clustering algorithm, distance measure, number of clusters criterion, number of clusters, evaluation of validity and evaluation of stability.

Single base commonsense / a priori segmentation (Concept 1)

Concept 1 segmentations require two steps: first, a splitting criterion is selected, typically by the marketing manager. Other indicators could come from prior literature as well as simulation. Each tourist is then classified as belonging to one of the segments based on their splitting criterion characteristic.

Second, tests are conducted to assess whether the resulting segments differ significantly from each other in relevant dimensions, such as travel benefits sought and vacation activities or sources of information used during the destination choice process . Such tests can either be conducted by independent Chi-square tests and ANOVAs, in which case the resulting p-values have to be corrected to ensure that the significance is not overstated. Alternatively, discriminant analyses or regression models can be used in which the dependent variable is membership to a segment and the independent variables are hypothesized to be significantly associated with being member of one of the segments.

The review of segmentation studies by consumer-based variables published in the last decade indicates that 56 per cent of segmentation studies in tourism use the Concept 1 approach. Thirty four per cent combine two segmentation bases to identify or construct market segments and only 10 per cent of studies use the pure single base data-driven approach.

Concept 1 studies vary in the splitting criterion chosen as well as in the sample size used to profile the segment of interest. The average sample size was 2080. The most frequently used splitting criteria (see Table 1) were behavioral (27 per cent) and

geographic (25 per cent), followed by socio-demographic (23 per cent) and psychographic (22 per cent). Within the dominant group of behavioral characteristics, vacation activities and frequency of visit were most popular.

[INSERT TABLE 1]

Note 1: A recent Concept 1 segmentation study

Kim et al. (2007) investigate gender differences in an online information search context. First, gender was used as a splitting criterion to divide the sample of 1334 respondents into groups of males and females. Second, groups were profiled and differences tested using ANOVA. Results indicate that women differ significantly from men in their online information search behavior: women tend to engage in a more exhaustive travel information search, use a wider variety of online and offline sources and have a higher frequency of visitation of websites. Kim et al.'s study illustrates how a simple Concept 1 segmentation can lead to valuable managerial insight, namely how to customize online information material to women and men.

Single base Data-driven / a posteriori / post-hoc segmentation (Concept 2)

When using Concept 2 approaches, the expectation is that a quantitative method will identify which market segments exist in the data. This expectation implies that natural segments exist in the data, as depicted in Figure 2a. Although revealing the existence of natural clusters was clearly the intention of the pioneers of clustering in the social sciences, market segmentation and market structure analysis (Aldenderfer & Blashfield, 1984; Frank, Massy & Wind, 1972; Myers & Tauber, 1977), segmentation researchers increasingly understand that survey data sets typically do not contain well separated density clusters. This has led to a foundational paradigm shift whereby market segmentation is now viewed as construction of artificial groupings which are managerially useful (Mazanec, 1997; Wedel and Kamakura, 1998). Typical segmentations are therefore represented by the data illustrations in Figure 2b and 2c as opposed to 2a.

This has major practical implications: working with data as depicted in 2c means that every computation aiming at grouping tourists will result in a different solution. Consequently, methodological decisions made during the computation impact on the resulting segments. Solutions may still be managerially useful, but the data analyst and the manager using the segmentation solution as the basis for long-term strategic decisions need to be aware that they are essentially choosing a "random" solution. Sometimes data contains pseudo-structure, as depicted in 2b. Although natural segments do not exist, pseudo-structure leads to segmentation solutions that can be reproduced across replications, thus providing more confidence in the reliability of the solution.

[INSERT FIGURE 2]

With this in mind, the following steps are taken when a data-driven segmentation study is conducted: the starting point is the selection of the segmentation base. Then a suitable grouping algorithm needs to be identified. Grouping algorithms are not objective tools which help researchers to see the structure of the data; instead they interact with data and have specific tendencies of constructing clusters of different shapes (Aldenderfer & Blashfield, 1984). The choice of algorithm affects the outcome. Everitt, Landau and Leese (2001) discuss the way in which a range of clustering algorithms interacts with the data.

Next the number of segments to be constructed has to be chosen. Choosing the best number of clusters is an unresolved problem (Thorndike, 1953). A number of approaches have been proposed and compared in the past (Milligan and Cooper, 1985; Dimitriadou et al., 2002), but no single best solution has emerged. The authors prefer the approach of computing a number of segmentation solutions for each number of clusters and comparing how often each pair of respondents is grouped together across replications, an approach proposed by Dolnicar, Grabler and Mazanec (1999).

Not that a large number of data-driven segmentation studies in tourism research use "factor cluster segmentation". This term describes a two-step approach whereby the original responses of tourists are factor analyzed, leading to a smaller subset of factors which are then used to cluster the data. This approach is not recommendable for market segmentation as demonstrated by Sheppard (1996) and Dolnicar and Grün (2008) because it does not help reveal heterogeneity based on consumer-based variables. These results are supported by a number of studies in management research and statistics (Aldenderfer and Blashfield, 1984; Arabie and Hubert, 1994; Ketchen and Shook, 1996, Milligan, 1996).

Concept 2 studies show fairly uniform trends. Sample sizes range from 100 to 1680, with an average of 498 respondents; the number of variables in the segmentation base range from five to 37 with an average of 17. Surprisingly, no association exists between sample size and the number of variables (Pearson correlation coefficient is -0.115). The studies with the largest number of variables in relation to the sample size only had three times as many respondents as variables. Fourteen per cent of studies have less than 10 times as many respondents as they use variables to identify homogeneous patterns. Formann's rule for latent class analysis can be used to provide an idea of reasonable sample sizes. Formann (1984) recommends - assuming binary question format - a sample of at least 2k to segment respondents using k variables; preferably 5*2k. According to this rule the studies reviewed would require, on average, 131,072 respondents. Although Formann's rule relates to parametric procedures, it provides guidance for exploratory market segmentation, as it highlights that the complexity of the grouping increases exponentially with more variables.

The segmentation base chosen by segmentation researchers in tourism was almost exclusively psychographic (64 per cent) or behavioral (21 per cent, see Table 2). Among the behavioral studies, vacation activities were (14 per cent of behavioral studies). Among psychographic studies, values were the most frequently used segmentation base (21 per cent). Other psychographic bases included perceptions, motives, attitudes, mood, expectations and benefits sought.

[INSERT TABLE 2]

More than half of the segmentation studies used ordinal data, raising the question of which distance measure was used, as distance measures for ordinal data are not readily available, yet they form the basic computation underlying clustering algorithms. Only 14 per cent of studies provide information about the distance measure used making an assessment of the quality of the segmentation solution difficult.

Data structure was not investigated by any of the Concept 2 studies reviewed, nor did any study include a measure of stability, indicating that researchers were either not aware of the structure before grouping or that they did not report data structure. In any case, it is difficult to assess the validity of a study if it is not clear whether segments were identified or artificially constructed. External validity, however, is tested frequently: two thirds of studies report differences between segments using ANOVA and Chi squared tests. Corrections for multiple testing are rarely computed.

Only few clustering algorithms are used : 36 per cent use k-means clustering and 21 per cent using hierarchical cluster analysis. The majority of studies used the raw data for computations, but 43 per cent used factor analysis, typically resulting in between three and ten factors which are subsequently used as data for clustering. This is concerning because the segmentation is actually performed in a transformed space if original variables are not used. The methods for choosing the best number of clusters range from not mentioning how this selection was made (21 per cent) to using heuristics and personal judgment.

Two key findings result from this review: (1) Concept 2 studies are not as frequent as anticipated, because many studies use a subset of tourists as a staring point. (2) There still is room for improvement with respect to the theoretical and methodological underpinnings of these studies. Many appear to use a process following a formula which is based on prior work, suggesting that researchers may not be fully aware of the impact of methodological choices which affect the outcome of the study.

Note 2: A recent Concept 2 segmentation study.

Dallen (2007) recently conducted a Concept 2 segmentation study investigating public attitudes towards a railway line. Complete linkage hierarchical cluster analysis was used to define psychographic segments. Nine ordinal attitudinal statements from 282 respondents served as segmentation base. The raw data was used for calculations. No information is provided about data structure investigations or the distance measure used. While not directly stated, it can be inferred that the decision about the number of clusters was based on heuristic procedures (Dallen 2007, p. 189). The clusters identified were described on the basis of additional characteristics. Although Chi square tests were mentioned, no significance tests are reported to allow the reader to assess the distinctiveness of resulting segments. Also, no evidence of stability is presented.

The final segmentation solution resulting from this study consists of five user segments; “Train Devotees”, “Infrequent Enthusiasts”, “Train Tolerators”, “Contented Car Users” and “Last Resort Riders”. Recommendations include targeting the “Infrequent Enthusiasts” by improving the train environment to take advantage of this segment’s already favorable attitude towards rail travel. In addition, “Train Tolerators” could be

convinced to continue using the train by showing how the mode is preferable to car travel with regards to congestion and parking.

Double base stepwise segmentation approaches (Concepts 3-7)

Double base segmentations are combinations of the two pure forms of market segmentation discussed above. Typically, they are conducted in a stepwise manner (Concepts 3-6) by first selecting a sub-set of tourists and then undertaking another grouping. Often the pre-selection of tourists is not declared as an intrinsic part of the segmentation process, thus leading to the possible misclassification of such approaches as purely data-driven (Concept 2). A few studies have also been published which use two segmentation approaches simultaneously (Concept 7). This can be achieved by choosing two splitting criteria or segmentation bases (gender and age), splitting the data set separately using both criteria and then cross-tabulating the results to arrive at combined segments.

Because double base segmentations are combinations of commonsense and data-driven segmentations, they involve the same steps as the pure Concept 1 and 2 segmentations consequently also being endangered by the same methodological problems. Our review of segmentation studies indicates that double-base stepwise segmentation studies are conducted in a very similar way to Concept 2 studies (see Table 3): the average sample size was 1180 with a average number of 27 variables used. One study used as many as 157 variables with only 850 respondents. More than two thirds used psychographic criteria (mostly traveler motives). All other studies used either behavioral variables (mostly vacation activities) or mixed variables. Most variables were ordinal in nature. Three studies could be identified which did investigate data structure prior to segmentation (Bloom, 2005; both studies in Dolnicar & Leisch, 2003).

[INSERT TABLE 3]

Factor analysis was the most frequently used method of pre-processing (60 per cent). Standardization of data occurred in only three studies, with the remainder of studies (34 per cent) not carrying out any pre-processing. K-means (42 per cent) and Ward's hierarchical clustering (17 per cent) or a combination of the two (9 per cent) represented the most popular algorithms. Of the 47 Concept 3-7 studies, only 13 stated the distance measure used. The primary method stated was squared Euclidean distance, a method not suitable for ordinal answer formats. Thirty six per cent did not report the method used to determine the number of clusters; when reported, personal judgment (26 per cent) or some sort of heuristic procedure (21 per cent) were more common. Concept 3-7 studies resulted in between two and seven segments. External validity was tested in 85 per cent of cases, mainly using ANOVAs (11 per cent), Chi squared tests (13 per cent), a combination of both (28 per cent) or discriminant analysis (15 per cent). Three of the Concept 3-7 studies evaluated the stability of results (two studies reported in Dolnicar and Leisch, 2003; Andreu et al., 2005).

Note 3: A recent Concept 5 segmentation study

Sung (2004) selected a subset of 892 adventure travelers and then conducted a data-driven segmentation using 26 socio-demographic, psychographic (perceptions) and behavioral variables (trip related factors including frequency of trip, expenditure, information sources) measured on different answer formats. Data structure was not investigated. Data was standardized before clustering, a reasonable decision given that variables with different scales were used. K-means clustering was the algorithm chosen. The number of clusters (6) was determined using information such as "distances between final cluster centers, iteration history, final cluster centers, number of cases in clusters and an ANOVA table" (Sung, 2004: 348).

Resulting segments were labeled "General Enthusiasts"; "Budget Youngsters"; "Soft Moderates"; "Upper High Naturalists"; "Family Vacationers" and "Active Soloists" and differed on factors such as household size, income and perceptions of adventure. Management recommendations were provided. For example, "General Enthusiasts" were found to be more likely to appreciate and participate in adventurous activities, whereas "Soft Moderates" were less likely to perceive adventure as highly. Therefore, differing marketing activities were recommended depending on the segment. The validity of this final cluster solution was investigated using discriminant analysis. The stability of the cluster solution was not investigated.

CONCLUSIONS AND IMPLICATIONS

The main results of this review indicate that segmentation research in tourism is still dominated by commonsense segmentation (Concept 1) studies. Such studies are sometimes viewed as "less sophisticated", but in fact represent the simplest model if tourism managers know which splitting criterion matters. Tests indicating how segments resulting from the split actually differ from each other should be provided to give tourism managers confidence that the splitting criterion does in fact lead to distinctly different segments.

Data-driven segmentation studies are heavily utilized in tourism research, both in their pure form (Concept 2) and as part of double-base segmentation studies (Concepts 4-6). Psychographic and behavioral variables are most frequently used as segmentation bases.

A number of methodological aspects have been identified which could be improved in future segmentation studies including a data-driven component.

(1) The number of variables should be reasonable given the sample size. Formanns' rule can be used as a guide.

(2) The structure of empirical data sets should be explored before clustering in order to be able to communicate to the end users whether naturally occurring segments were revealed, stable artificial clusters were constructed or unstable artificial clusters are presented.

(3) Data should not be factor analyzed before clustering. Instead, the raw data should be used where possible.

(4) Cluster solutions should be validated using both stability and external validity measures.

(5) The detailed parameters of the computations should be reported, including the precise algorithm chosen, distance measure used and the approach taken for choosing the number of clusters.

(6) Finally, it would be advantageous if segmentation studies would take into consideration positioning and competition issues. Currently they are treated as stand-alone components of strategic marketing, which does not mirror the reality that tourism destinations and businesses face.

Surprisingly, the picture that emerges from this review of segmentation studies published in the last decade mirrors previous findings of review studies. This is surprising because previous reviews have made explicit recommendations about how segmentation studies could be improved in the future. Specifically, Sheppard (1996) emphasized the need for caution when selecting factor analysis as a means of pre-processing, asserting that cluster analysis on raw item scores produces more accurate data-driven results.

Frochot and Morrison (2000) recommended selecting the most discriminating criteria for inclusion into the segmentation base, as practiced by Gitelson and Kerstetter (1990) who eliminated variables that were considered unimportant by more than 90 per cent of respondents. This approach helps researchers reduce the number of variables without conducting factor analysis. Dolnicar (2002) recommended that the number of variables should be carefully selected relative to the sample size and that stability should be evaluated for obtaining a reliable solution. The issue of stability was raised by Dolnicar and Leisch (2003) who presented a method (bagged clustering) that automatically accounts for stability and by Frochot and Morrison (2000) who recommended repeated measurements as the method of choice. Dolnicar, Mazanec and Graber (1999) pointed out the importance of strategic integration of market segmentation, product positioning and competition aspects.

It can be concluded that tourism research has come a long way with respect to acknowledging that consumers are not one homogenous mass which can be expected to enjoy the same tourism product. A large number of segmentation studies have been conducted over the past decade which acknowledge differences between consumers and attempt to learn how to best match consumer needs of specific sub-segments. There is, however, still significant room for improvement.

REFERENCES

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

Andreu, L., Kozak, M., Avci, N., & Cifter, N. (2005) 'Market segmentation by motivations to travel: British tourists visiting Turkey', *Journal of Travel and Tourism Marketing, 19*(1), 1-14.

Arabie, P., & Hubert, L. (1994) 'Cluster analysis in marketing research', in R. Bagozzi (Ed.), *Advanced Methods of Marketing Research*. Blackwell: Cambridge.

Bloom, J. (2005) 'Market segmentation, a neural network application', *Annals of Tourism Research, 32*(1), 93-111.

Buchta, C., Dolnicar, S., & Reutterer, T. (2000) '*A nonparametric approach to perceptions-based market segmentation*', (Vol. 2). Berlin: Springer.

Dallen, J. (2007) 'Sustainable transport, market segmentation and tourism: The Looe Valley branch line railway, Cornwall, UK', *Journal of Sustainable Tourism, 15*(2), 180-199.

Dimitriadou, E., Dolnicar, S., & Weingessel, A. (2002) 'An examination of indexes for determining the number of clusters in binary data sets' *Psychometrika, 67*(1), 137-160.

Dolnicar, S. (2002) 'Review of data-driven market segmentation in tourism', *Journal of Travel and Tourism Marketing, 12*(1), 1-22.

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. & Grün, B. (2008) 'Challenging "Factor Cluster Segmentation', *Journal of Travel Research, 47, 63-71.*

Dolnicar, S., Grabler, K., & Mazanec, J. A. (1999) 'A tale of three cities: Perceptual charting for analyzing destination images', in A. Woodside (Ed.), *Consumer Psychology of Tourism, Hospitality and Leisure* (pp. 39-62). New York: CAB International.

Dolnicar, S., & Leisch, F. (2001, 5-8 July). *Knowing what you get - A conceptual clustering framework for increased transparency of market segmentation studies.* Paper presented at the Marketing Science, Wiesbaden, Germany.

Dolnicar, S., & Leisch, F. (2003) 'Winter tourist segments in Austria – Identifying stable vacation styles for target marketing action', *Journal of Travel Research, 41*(3), 181-193.

Dolnicar, S., & Mazanec, J. A. (2000), 'Holiday styles and tourist types: Emerging new concepts and methodology', in W. C. Gartner & D. W. Lime (Eds.), *Trends in Outdoor Recreation, Leisure 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, R. E., Massy, W. F., & 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.

Gitelson, R. J., & Kerstetter, D. L. (1990) 'The relationship between sociodemographic variables, benefits sought and subsequent vacation behavior: A case study', *Journal of Travel Research, 28*(3), 24-29.

Haley, R. J. (1968) 'Benefit segmentation: A decision oriented research tool', *Journal of Marketing, 32*(30), 35.

Ketchen, D. J. j., & Shook, C. L. (1996) 'The application of cluster analysis in strategic management research: an analysis and critique', *Strategic Management Journal, 17*, 441-458.

Kim, D., Jehto, X. Y., & Morrison, A. M. (2007) 'Gender differences in online travel information search: Implications for marketing communications on the internet', *Tourism Management, 28*, 367-378.

Mazanec, J. A. (Ed.). (1997) *International city tourism: Analysis and strategy*. London: Pinter/Cassell.

Mazanec, J. A. (2000) 'Market segmentation', in J. Jafari (Ed.), *Encylcopedia of Tourism*. London: Routledge.

Mazanec, J. A., & Strasser, H. (2000), *A nonparametric approach to perceptions-based market segmentation: Foundations.* Berlin: Springer.

Mazanec, J. A., & Strasser, H. (2007), 'Perceptions-based analysis of tourism products and service providers', *Journal of Travel Research, 45*(4), 387-401.

McCleary, K. W., Weaver, P. A., & Hsu, C. H. (2006) 'The relationship between international leisure travellers' origin country and product satisfaction, value service quality, and intent to return', *Journal of Travel and Tourism Marketing, 21*(2/3), 117-130.

Milligan, G. W. (1996) 'Clustering validation: Results and implications for applied analyses', in P. Arabie & L. Hubert (Eds.), *Clustering and Classification*. River Edge: World Scientific Publications.

Milligan, G. W., & Cooper, M. C. (1985). 'An examination of procedures for determining the number of clusters in a data set', *Psychometrika, 50*(2), 159-179.

Myers, J. H., & Tauber, E. (1977) *Market Structure Analysis*, Chicago: American Marketing Association.

Reece, W. (2004) 'Are senior leisure travellers different?', *Journal of Travel Research, 43*(1), 11-18.

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).

Sung, H. H. (2004) 'Classification of adventure travellers: Behavior, decision making and target markets', *Journal of Travel Research, 42*(4), 343-356.

Thorndike, R. L. (1953) 'Who belongs in the family?', *Psychometrika, 18*(4), 267-276.

Wedel, M., & Kamakura, W. (1998). *Market segmentation: Conceptual and methodological foundations*. Boston: Kluwer Academic Publishers.

TABLES AND FIGURES

[List of Captions]

Figure 1: Segmentation approaches based on methods used for grouping

Figure 2: Examples of three data structures underlying a two-dimensional segmentation task (modified from Dolnicar & Leisch, 2001)

Table 1: Segmentation variables used in Concept 1 studies

Table 2: Methodological characteristics of Concept 2 studies

Table 3: Methodological characteristics of double base segmentation studies

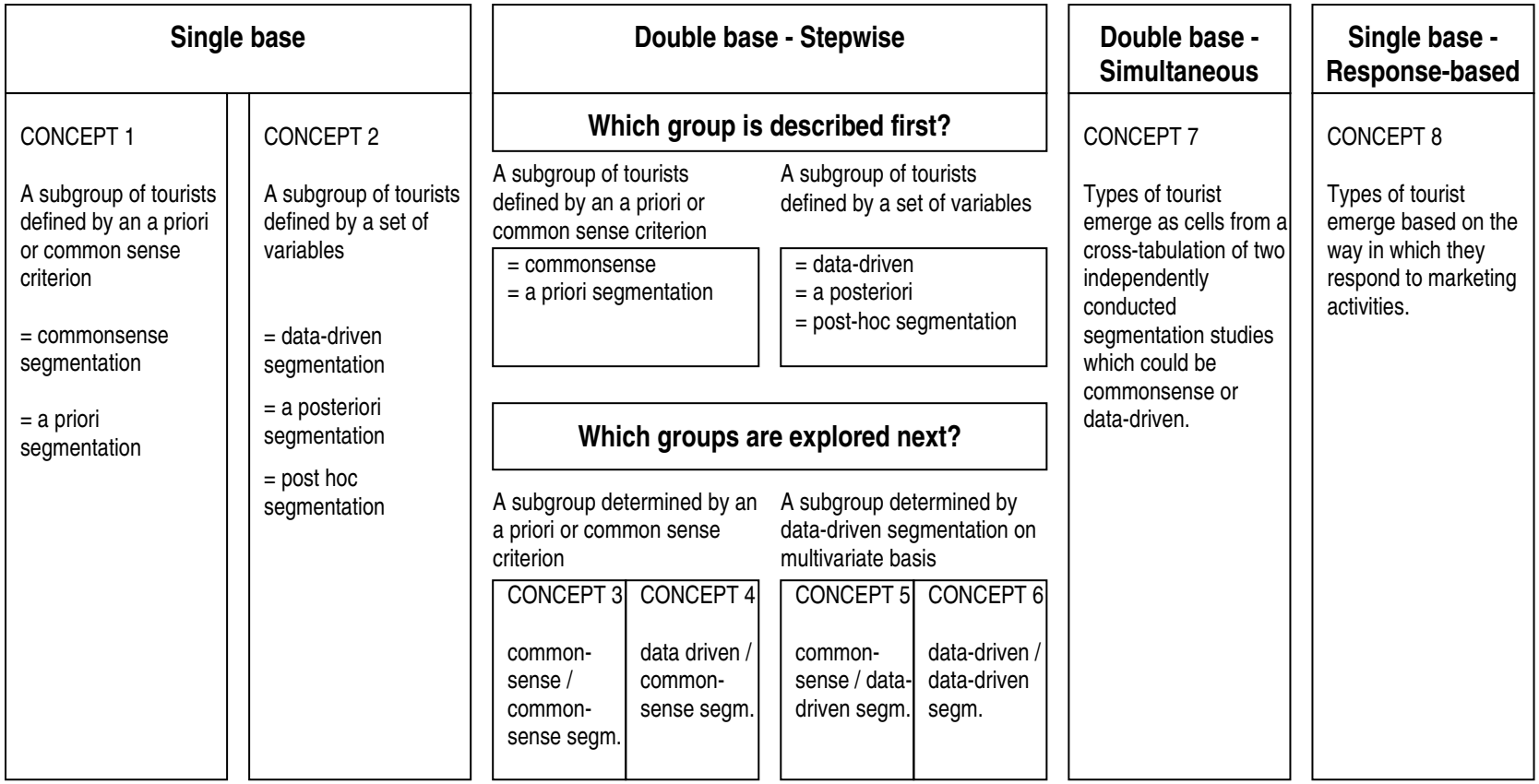
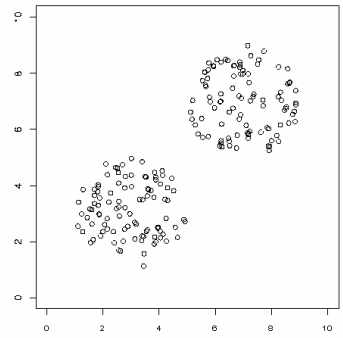


Figure 1: Segmentation approaches based on methods used for grouping

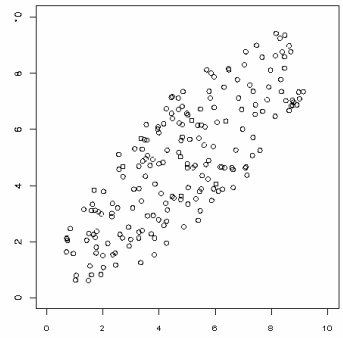
Figure 2

Examples of three data structures underlying a two-dimensional segmentation task (modified from Dolnicar & Leisch, 2001)

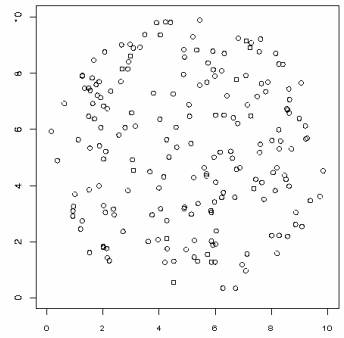
2a: True segment structure



2b: Pseudo structure



2c: No structure



Note: Figure 2 assumes only a two-dimensional segmentation problem, meaning that only two consumer-based variables are used as segmentation base. In reality data-driven segmentation studies use significantly more variables than two.

Table 1

Segmentation variables used in Concept 1 studies

Variable		Frequency	Per cent
Segmentation base			
	Behavioral	**21**	**27%**
	Activities undertaken	8	10%
	Frequency of visit	8	10%
	Expenditure	5	7%
	Geographic	**19**	**25%**
	Socio-demographic	**18**	**23%**
	Gender	7	9%
	Age	4	5%
	Lifecycle Stage	2	3%
	Disability	2	3%
	Other Socio-demographic	3	4%
	Psychographic	**17**	**22%**
	Values	5	7%
	Motives	3	4%
	Attitudes	3	4%
	Expectations	2	3%
	Other Psychographic	4	5%
	Mixed / Other	**2**	**3%**

Table 2: Methodological characteristics of Concept 2 studies

Variable		Frequency	Per cent
Segmentation Base Used	**Psychographic**	**9**	**64%**
	Values	3	21%
	Other	6	43%
	Behavioral	**3**	**21%**
	Activities undertaken	2	14%
	Search behavior	1	7%
	Mixed Bases	**2**	**14%**
Data format of variables	Ordinal	9	64%
	Binary	1	7%
	Metric	1	7%
	Mixed	1	7%
	Not stated	2	14%
Evaluation of data structure	Yes	0	0%
	No	14	100%
Method of pre-processing	No pre-processing	8	57%
	Factor analysis	6	43%
Number of factors identified	3	1	17%
	4	3	50%
	5	1	17%
	10	1	17%
Clustering algorithm chosen	K-means	5	36%
	Hierarchical (not further specified)	3	21%
	Other	6	43%
Distance measure	Stated	2	14%
	Not stated	12	86%
Method for determining the number of clusters	Personal judgment	5	36%
	Heuristic procedure	3	21%
	Dendrogram from hierarch. step	3	21%
	Not stated	3	21%
Number of segments (clusters) selected	2	2	14%
	3	2	14%
	4	5	35%
	5	2	14%
	7	2	14%
	10	1	7%
Evaluation of validity	ANOVA	3	21%
	ANOVA and Chi square Combination	3	21%
	MANOVA	1	7%
	Discriminant analysis	1	7%
	Mixture of the above	1	7%

	Not evaluated	5	35%
Evaluation of stability	Yes	0	0%
	No	14	100%

Table 3: Methodological characteristics of double base segmentation studies

Variable		Frequency	Per cent
Segmentation Base Used	**Psychographic**	**34**	**72%**
	Motives	17	36%
	Benefits sought	5	11%
	Perceptions	4	9%
	Preferences	2	4%
	Attitudes	2	4%
	Other psychographic	4	9%
	Behavioral	**9**	**19%**
	Activities undertaken	7	15%
	Other behavioral	2	4%
	Mixed Bases	**4**	**9%**
Data format of variables	Ordinal	33	70%
	Binary	3	6%
	Metric	2	4%
	Nominal	1	2%
	Mixed	4	9%
	Not stated	4	9%
Evaluation of data structure	Yes	3	6%
	No	44	94%
Method of pre-processing	No pre-processing	16	34%
	Factor analysis	28	60%
	Standardization	3	6%
Number of factors identified	3	3	10%
	4	6	21%
	5	11	38%
	6	3	10%
	8	4	14%
	10	1	3%
	34	1	3%
Clustering algorithm chosen	k means	20	42%
	Ward's	8	17%
	Combination of k means and Wards	4	9%
	Hierarchical (not further specified)	4	9%
	Other	5	11%
	Not stated	5	11%
Distance measure	Stated	13	28%
	Not stated	38	72%
Method for determining the number of clusters	Personal judgment	12	26%
	Heuristic procedure	10	21%
	Dendrogram from hierarch. step	6	13%
	Other	2	4%

	Not stated	17	36%
Number of segments (clusters) selected	2	6	13%
	3	11	24%
	4	15	32%
	5	9	19%
	6	5	9%
	7	1	2%
Evaluation of validity	ANOVA and Chi Square Combination	13	28%
	Discriminant Analysis	7	15%
	Chi Squared testing	6	13%
	ANOVA	5	11%
	Comparison with external variables	3	6%
	None	7	15%
	Mixture of the above	6	13%
Evaluation of stability	Yes	3	6%
	No	44	94%