2004

Improved understanding of tourists' needs – cross-classification for validation of data-driven segments

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

Improved understanding of tourists' needs – cross-classification for validation of data-driven segments

Abstract

Data-driven segmentation has become standard practice in strategic marketing. Typically, however, respondents are grouped only once, implicitly assuming deterministic nature of the segmentation methods applied. Once the segments are derived, background variables are used to test the significance of the difference between clusters indicating external validity of the market segments. High external validity implies a high level of trustworthiness of the solution and thus managerially useful market segments to choose from. However, single runs of explorative analysis remain only a weak basis for good long-term managerial decisions. In this study a different approach is suggested to improve the quality of the market structure insight for decision-making: two data-driven segmentation solutions are constructed independently. Association between them is used as an additional internal validity indicator. Benefit and behavioural segmentation bases are used to illustrate the concept: surfers provided information regarding the importance of aspects of proposed surf destinations and destinations they had previously visited. Segments resulting from both bases are profiled: they notably differ in background information. Significant association between these solutions supports the validity and managerial usefulness for target segment choice for the following targeted marketing action. Using this procedure results in the identification of stable consumer groups, which in turn leads to an improved understanding of customer segments; and thus enables the industry to improve the quality by customizing the product.

Keywords

segmentation, validation, post-hoc (a posteriori, data-driven)

Disciplines

Business | Social and Behavioral Sciences

Publication Details

Dolnicar, S, Improved understanding of tourists' needs: cross-classification for validation of data-driven segments, in Thyne, M and Laws E. (eds.), Hospitality, Tourism and Lifestyle Concepts: Implications for Quality Management and Customers Satisfaction, 2004, 141-156.

Improved understanding of tourists' needs – cross-classification for validation of data-driven segments

Sara Dolnicar

School of Management, Marketing & Employment Relations

University of Wollongong

Northfields Ave, Wollongong, NSW 2522, Australia

sara_dolnicar@uow.edu.au, Phone: 61 +2 4221 3862

Acknowledgement

I thank Jess Pointing, Martin Fluker and The Surf Travel Company for providing the interesting empirical data set that allows illustrating the validation concept proposed.

Improved understanding of tourists' needs – cross-classification for validation of data-driven segments

Abstract

Data-driven segmentation has become standard practice in strategic marketing. Typically, however, respondents are grouped only once, implicitly assuming deterministic nature of the segmentation methods applied. Once the segments are derived, background variables are used to test the significance of the difference between clusters indicating external validity of the market segments. High external validity implies a high level of trustworthiness of the solution and thus managerially useful market segments to choose from. However, single runs of explorative analysis remain only a weak basis for good long-term managerial decisions. In this study a different approach is suggested to improve the quality of the market structure insight for decision-making: two data-driven segmentation solutions are constructed independently. Association between them is used as an additional internal validity indicator.

Benefit and behavioural segmentation bases are used to illustrate the concept: surfers provided information regarding the importance of aspects of proposed surf destinations and destinations they had previously visited. Segments resulting from both bases are profiled: they notably differ in background information. Significant association between these solutions supports the validity and managerial usefulness for target segment choice for the following targeted marketing action.

Using this procedure results in the identification of stable consumer groups, which in turn leads to an improved understanding of customer segments; and thus enables the industry to improve the quality by customizing the product.

Keywords: segmentation, validation, post-hoc (a posteriori, data-driven) segmentation

Introduction

It has become a standard procedure to account for consumer heterogeneity in the market place by grouping individuals based on the similarity of their answers given in a survey. Responsible for this might be:

- Socio-demographic criteria. For instance, Dodd & Bigotte (1997) group winery visitors according to demographics.
- Behavioural information. An example is provided by Hsu & Lee (2002) who construct segments based on motor coach selection criteria.
- Benefit variables. This concept was first introduced by Haley in 1968 and has rapidly risen in popularity; a recent example is provided by Silverberg, Backman & Backman (1996) who investigate benefit segments among nature-based tourists.

The procedure of investigating existing or constructing artificial but managerially useful groups of similar customers is referred to as *a posteriori* (Mazanec, 2000), *post-hoc* (Wedel & Kamakura, 1998), or data-driven market (Dolnicar, 2002a) segmentation. Homogeneous market segments derived in such way can be subsequently targeted in a more efficient manner (Frank, Massy & Wind, 1972; Wedel & Kamakura, 1998): sub-group's demands can be met better.

Typically, market segments are derived by clustering (determining sub-groups by computing the similarities of consumers) the data only once. Detailed reviews of emerged standards in data-driven market segmentation were investigated by Baumann (2000) and Dolnicar (2002). In general, very little effort is made to validate the managerial usefulness of segments. Accepted ways of validating results include the use of repetitive algorithms, manual repetition, or testing of additionally available information. The latter approach can often be found in segmentation studies published in academic journals. In the present article, an additional approach for validation of data-driven segmentation results is illustrated which makes use of multiple segmentation bases existing in one data set. After independent data-driven grouping, the association of the solutions is used as an indicator of validity.

The procedure proposed has been used prior to this study for the purpose of identifying integrated vacation styles (Dolnicar & Mazanec, 2000; Dolnicar & Leisch, 2003). Here this approach is presented as a useful additional tool for evaluating the internal validity of data-driven market segments. By evaluating the validity of segmentation results, the customer segments chosen as targets are tested more thoroughly; as such they form a stronger basis for understanding the needs of defined customer groups. Better understanding of the customers enables the hospitality and tourism industry to better customize its products and consequently better match the customers' needs and desires. This is expected to lead to higher quality products optimally harmonized with demand for the particular target segments chosen, higher consumer satisfaction and increased loyalty.

Methodology

The proposed validation procedure is illustrated using an empirical example. The data set includes 430 respondents (active surfers) to an online-survey. Two pieces of information were used for independent construction of segments: (1) the importance attached to different

destination factors (benefits tourists seek when choosing a particular destination), and (2) determination of prior visits to 30 international surf destinations. The latter consists of Yes or No statements. The importance block includes 17 items indicating importance of factors for destination choice. In addition, the survey includes other questions relating to surf behaviour, personal characteristics and general travel behaviour, the answers to which can be used for profiling and determining significant differences between segments.

For the construction of the surfer segments self-organising neural network procedures were applied: self organising feature maps (SOFMs, Kohonen, 1984) and topology representing networks (TRNs, Martinetz & Schulten, 1994). The term *neural network* refers to techniques from the area of artificial intelligence that can be used for the purpose of data analysis. Self-organising neural network procedures – meaning that they "learn" without being given feedback about the correct answer - were first introduced into the area of market segmentation by Mazanec (1992), their main advantage being additional topological market structure information resulting from only one step of partitioning analysis. The benefit segments based on importance statements were built using the less rigid TRN approach, whereas the behavioural segments were forced into a SOFM grid, in order to reveal possible topological neighbourhood patterns using a framework that allows simple definition of similarity of segments. For both networks random starting points were chosen and training was allowed for 90 epochs with a decreasing learning rate. The software used is freely available from the Vienna University of Economics and Business Administration (http://charly.wu-wien.ac.at/software/).

Behavioural segmentation

Six segments were chosen, because this solution rendered sufficiently large and yet well-profiled segments[1]. The resulting solution is described in detail with regard to the implications for surf tourism marketing in Dolničar & Fluker (forthcoming) and is outlined in Figure 1. Each bar chart represents one segment. The bars indicate the proportion of segment members stating that they had visited each destination, and the line represents the sample average. The resulting segments are surprisingly distinct. Behavioural segment 1 (Indonesia segment) has a very strong focus on Indonesia as a surf destination and includes roughly 10 % of the respondents[2], B2-members (America segment, 24 %) are above average in surfing American destinations. Segment B3 (Western Australia and Indonesia segment, 8 %) is characterized by a combination of Western Australian and Indonesian destinations, B4 (Australia segment, 16 %) represents a group of surfers that almost exclusively surfs at the coasts of Australia (the only other destinations mentioned, on average, more often was the Philippines). Surfers assigned to B5 (all destinations, 17 %) declare that they have surfed anywhere in the world more often than the average. Here, answer tendencies might distort results; the segment should be interpreted with care. Finally, B6-surfers (North Eastern Australia, 25 %) have so far surfed in Queensland and at the north coast of New South Wales; they represent a second group of Australia-surfers.

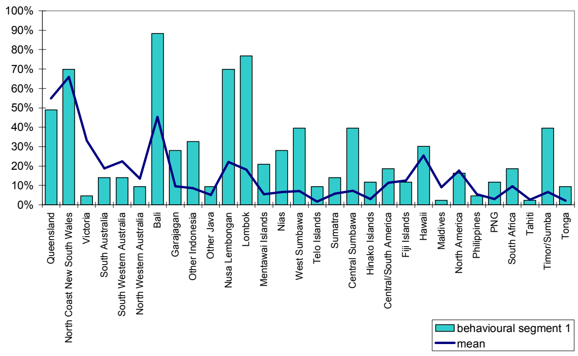
---

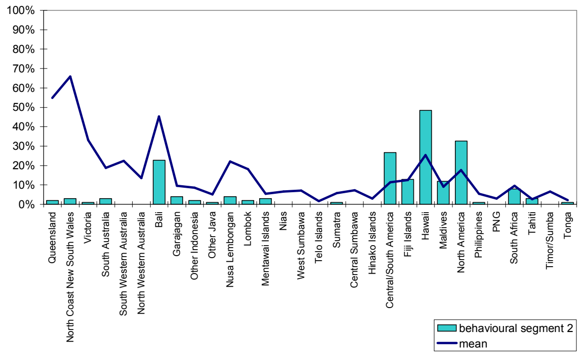
[1] The author is aware of the fact that every grouping of 430 respondents in a thirty dimensional space is a very rough partitioning and that the sample size does represent a major limitation. However, the aim of this article is to illustrate a simple internal validity measure, not to gain insight into surf tourist. In case the latter were the aim, a larger and representative sample size would be needed.

[2] These percentages are only rough because the sample was not drawn following a representative sampling procedure. Strictly speaking the percentages are proportions of the sample only, not the population.

In addition to the segmentation base, descriptive information was available in the data set. Such information is used to further describe the segments and to investigate whether the grouping chosen actually represents distinct groups.

A number of significant differences between the segments are revealed: The average age varies (ANOVA p-value = 0.000) from 27 to 33 years, with surfers focusing on American sites representing the oldest group. The years of surfing experience (Chi square p-value = 0.000, this p-value is Bonferroni-corrected to adjust for the fact that multiple tests are computed on the basis of the same data set) distinguish the behavioural segments: again, the America-surfers apparently have most experience, whereas the surfers visiting Indonesia and Western Australia as well as the NSW/Queensland group are least experienced. With regard to the length of stay (0.000), Indonesia-surfers stay longest, with 23 % stating that they stay for 5-8 weeks. The America-surfers (B2), the Indonesia and Western Australia segment (B3) as well as the Eastern Australia group (B6) have the shortest lengths of stay with about two thirds staying less than two weeks.





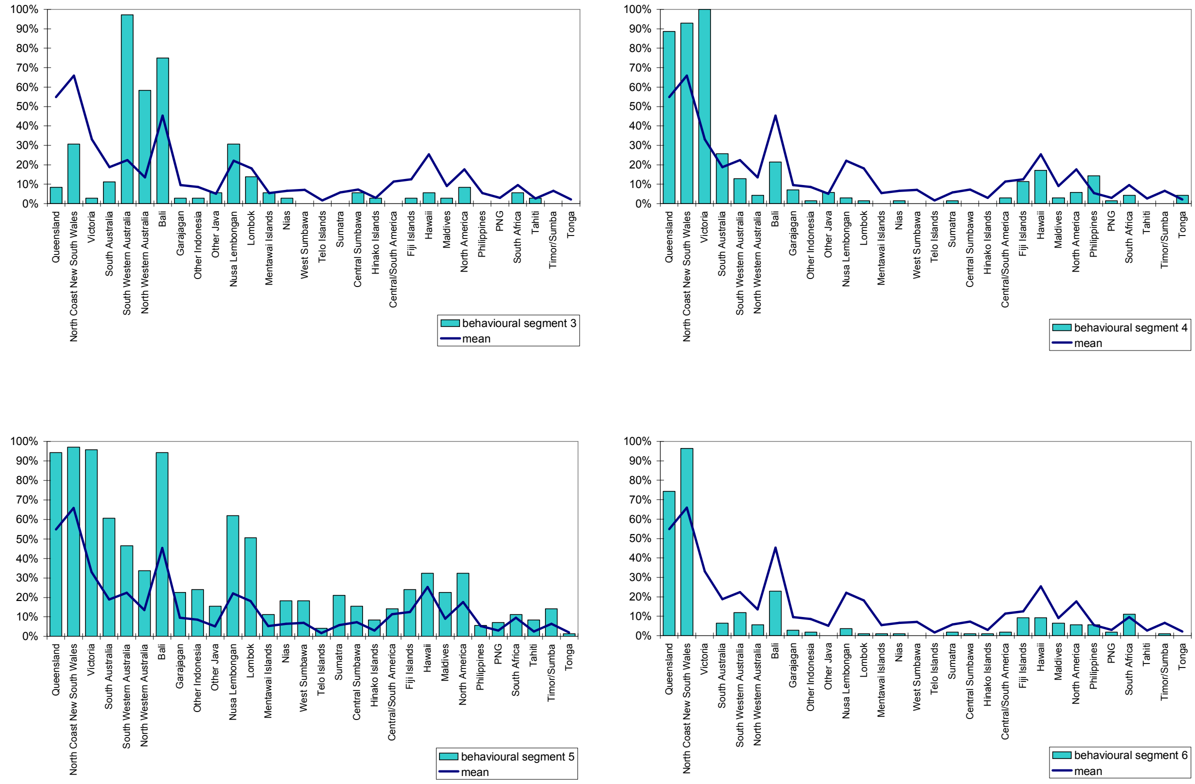


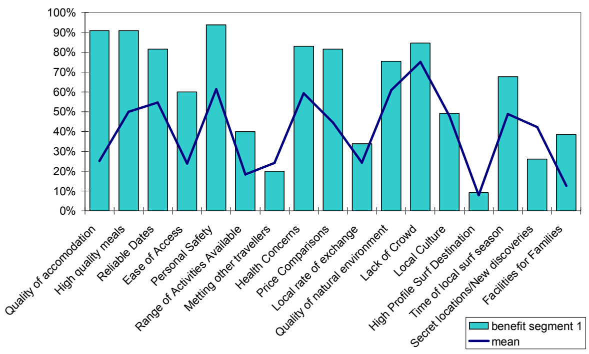
Figure 1: Behavioural segments (Source: Dolničar & Fluker, forthcoming)

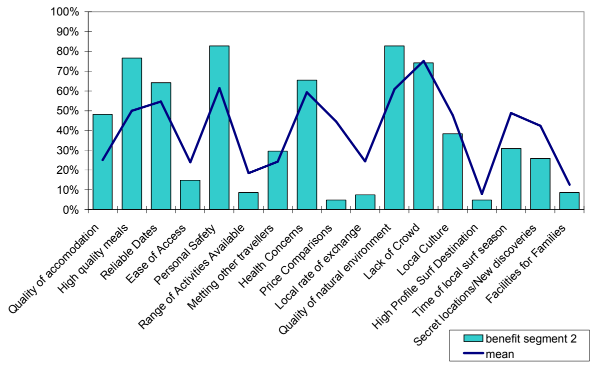
Further significant distinguishing criteria of defining segments include the preferred wave type (0.004), the regularity of undertaking surf trips (0.001), the interest in destination novelty (0.048), education level (0.010) and income (0.006) although no significant differences in daily budget are detected.

Benefit segmentation

A five-cluster solution was chosen because contingency tables between solutions with different numbers of clusters reveal that there is a high congruence of surfer types with both the four and the six-segment solution indicating local stability of types over solutions with a different number of segments. Figure 2 outlines the segments derived (they were discussed in detail with regard to tourism marketing implications in Dolnicar & Fluker, 2003). Segment 5 (P5, "radical adventurers", 19 %) represents the most distinct group of surfers: the time of the

local surfing season as well as secret locations are important to the majority of this segment, local culture, the lack of crowd and quality of natural environment also play a central role. P4 (23 %, not given in Figure 2) does not identify anything to be important at all[3]. Segment P2 ("luxury surfers", 19 %) is interesting from the viewpoint that neither price nor exchange rate are of particular interest. It is more important to this group that accommodation is good, food is excellent and safety is assured. Segments P1 and P3 are very similar to each other. The surfing-related items are important to these segments, personal safety and health play an important role, high-quality meals and reliable dates are appreciated. The main differentiating factors are the availability of facilities for families and the quality of accommodation in P1 ("price-conscious safety seekers", 15 %) and the search for new locations and discoveries as well as the lack of crowd in P3 ("price-conscious adventurers", 24 %).





[3] This is a typical phenomenon that occurs in data-driven market segmentation: one or two segments function as collecting points for answer patterns that are either above or below average with regard to all variables. Such segments profiles are ambiguous, because they could either represent a real answer pattern that can be interpreted (in this case, for instance, that none of the tourists see any benefit in any of those aspects) or an answer pattern that should not be interpreted, because it might result from respondent fatigue or lack of motivation to complete the questionnaire properly.

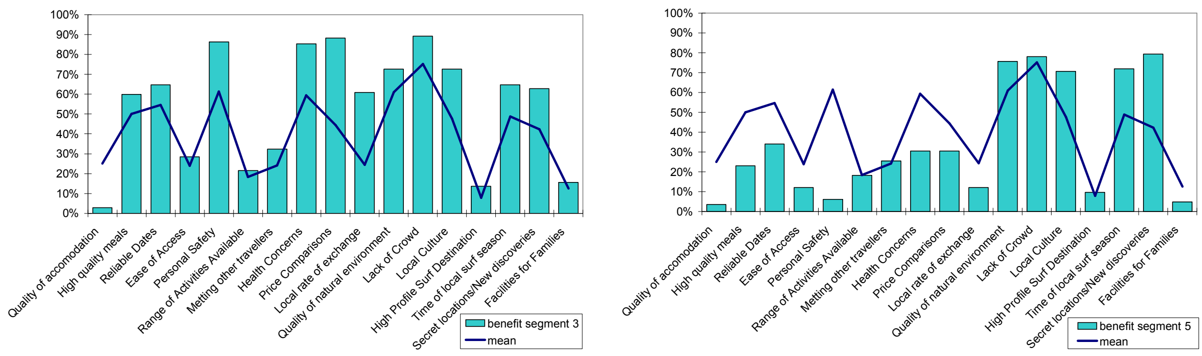


Figure 2: Benefit surfer segments (Source: Dolničar & Fluker, 2003)

The analysis of descriptive variables indicates significant discrimination between these segments. Radical adventurers (P5) are the youngest group; the price-conscious safety seekers (P1) represent the oldest group (ANOVA p-value = 0.000). This is well mirrored in the years of surfing experience (Bonferroni-corrected Chi square p-value = 0.009), where 37 % of the latter claim that they have been surfing for more than 20 years. Regarding surfing ability (0.001), very few members of each segment dare to call themselves “highly advanced”, but half of both the price-conscious safety seekers (P2) and the price-conscious adventurers (P3) regard themselves as “advanced”. Also of interest: the highest number of “beginner” classifications appears in group P3. Further significant differences include the preferred wave size (0.010), the preferred wave type (0.009), movement during the stay (0.004), education (0.000) and income level (0.000).

Association between solutions

The contingency table of memberships for both segmentation solutions is presented in Table 1. This is the core component of the proposed validation procedure. The assumption underlying this internal validity investigation is that segments of consumers that truly exist in the market can be repeatedly revealed from different perspectives, looking at different tourist characteristics.

The association is tested using the Pearson Chi square test that renders a p-value of 0.022, thus indicating significant interdependence between the two solutions at the 95 % level (Cramer's V is significant with a p-value of 0.022 as well and amounts to 0.142). The values for the segments conspicuous of being (at least partially) answer tendencies are given in grey. The cells in which the observed counts strongly exceed the expected number of respondents appear in boldface; cells in which the observed counts are lower than the expected value are marked in italic. For each cell the observed counts, the expected cell frequency and the percentages with regard to both segmentations are provided. Managerially, the boldface cells are of interest as they point to the market segments that are identified both from the benefit as well as from the behavioural perspective. Surfers are consequently significantly more often than one would expect located in both the behavioural and benefit segment represented by the boldface cross-tabulation cell.

Table 1: Interdependence of benefit and behavioural segments

		P1 price-conscious safety seeker	P2 luxury surfer	P3 price-conscious adventurers	P4	P5 radical adventurer
B1 Indonesia	Observed	4	*2*	10	13	**14**
	Expected	6,5	*8,1*	10,2	10,0	**8,2**
	% of behavioural	9,3	*4,7*	23,3	30,2	**32,6**
	% of benefit	6,2	*2,5*	9,8	13,0	**17,1**
B2 America	Observed	**22**	18	27	24	*10*
	Expected	**15,3**	19,0	24,0	23,5	*19,3*
	% of behavioural	**21,8**	17,8	26,7	23,8	*9,9*
	% of benefit	**33,8**	22,2	26,5	24,0	*12,2*
B3 Western Australia and	Observed	5	*3*	8	9	**11**
	Expected	5,4	*6,8*	8,5	8,4	**6,9**
	% of behavioural	13,9	*8,3*	22,2	25,0	**30,6**

Indonesia	% of benefit	7,7	*3,7*	7,8	9,0	**13,4**
B4	Observed	8	**16**	17	*10*	**19**
Eastern Australia	Expected	10,6	**13,2**	16,6	*16,3*	**13,3**
	% of behavioural	11,4	**22,9**	24,3	*14,3*	**27,1**
	% of benefit	12,3	**19,8**	16,7	*10,0*	**23,2**
B5	Observed	10	14	18	16	13
All destinations	Expected	10,7	13,4	16,8	16,5	13,5
	% of behavioural	14,1	19,7	25,4	22,5	18,3
	% of benefit	15,4	17,3	17,6	16,0	15,9
B6	Observed	16	**28**	*22*	28	*15*
North Eastern Australia	Expected	16,5	**20,5**	*25,9*	25,3	*20,8*
	% of behavioural	14,7	**25,7**	*20,2*	25,7	*13,8*
	% of benefit	24,6	**34,6**	*21,6*	28,0	*18,3*

Including one third of the members of the price-conscious safety seekers (P1) this group is highly over-represented in the behavioural segment that surfs in American waters. The luxury surfers (P2) prefer Australia: 23 % are assigned to B4 (Eastern Australia) and, 26 % to B6 (North Eastern Australia), but are under-represented in the Indonesia-segment. A slight deviation from the expected values occurs for the price-conscious adventurers (P3): they are under-represented among the surfers that include Queensland and New South Wales in their portfolio of surfing destinations (B6). Finally, the radical adventurers show the most profiled distribution: These surfers are more often than expected found in behavioural segments B1, B3 and B4 where the two first segments represent the Indonesia-region on the topological map and B4 the East of Australia. On the other hand, the less radical adventurers can be found among the members of B2, the America-centred segment, and B6, the group focusing on Queensland and New South Wales when going for a surf vacation. Those segments thus represent good opportunities for managers of surfing destinations because they are independently revealed both from the behavioural and the benefit side. The internal validity of

such segments is higher than it is the case for the remaining surfer groups. This gives management more confidence that the chosen target segment is not a random result or a methodological artefact.

Conclusions, limitations and future work

Two data-based market segmentations of surf tourists were cross-tabulated for the purpose of internal validation of derived market segments. One was constructed on the basis of benefits sought and the other on the basis of surfing destinations visited in the past. Both segmentation solutions render distinct segments significantly differing in various background variables. Significant association is detected between the two groupings that indicate systematic association between the benefit segments and the behavioural segments. This ads validity to the solution (besides the external substantiation obtained by testing of significant differences in descriptive information) and supports the assumption that segments that are homogeneous with regard to the importance respondents attach to different destination factors when choosing among them does interact with their actual choice of destinations.

The proposed procedure assists hospitality and tourism managers to avoid selecting market segments that are not backed by both behavioural and benefit characteristics. Australia as a surfing destination, for instance, is lucky to find that the market segment named “luxury surfers” likes to surf at Eastern and North Eastern Australian destinations. This segment is not very price-sensitive but the quality of accommodation and food is important for these tourists, and safety must be assured. Realising these facts enables the destination management to provide the expected kind of service and product thus increasing the quality of the service.

This study is only the first step toward investigating the usefulness of cross-classification for the validation of data-driven segment solutions. The limitation lies in the requirement to have

at least two question blocks of different nature in the guest survey to enable independent segment construction. In case of this particular illustration based on the surfer data set, the limitations are that only one single data set was studied that was not optimal due to the data collection method. The sample size was too small for a managerially relevant segmentation study given the large number of variables. The representativeness could not be assumed given the online survey technique used. Future work must include multiple replications of this approach with numerous data sets of varying nature and dimensionality. Analysis with artificial data could give guideline values for the various cluster combinations studied.

References

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

Dodd, T., & Bigotte, V. (1997). Perceptual Differences among Visitor Groups to Wineries. *Journal of Travel Research*, 35(3), 46-51.

Dolničar, S. (2002a). Review of Data-Driven Market Segmentation in Tourism. *Journal of Travel & Tourism Marketing.* 12(1), 1-22.

Dolničar, S. (2002b). A Review of Unquestioned Standards in using Cluster Analysis for Data-driven Market Segmentation. *CD Proceedings of the Australian and New Zealand Marketing Academy Conference (ANZMAC).*

Dolnicar, S., & Fluker, M., (2003). Who's Riding the Wave? An Investigation Into Demographic and Psychographic Characteristics of Surf Tourists. *CD proceedings of the 13th International Research Conference of the Council for Australian University Tourism and Hospitality Education (CAUTHE).*

Dolničar, S., & Fluker, M. (forthcoming). Behavioural Market Segments Among Surf Tourists – Investigating Past Destination Choice. *Journal of Sports Tourism.*

Dolničar, 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., & Mazanec, J. (2000). Holiday Styles and Tourist Types: Emerging New Concepts and Methodology, in: Gartner, W.C., and Lime, D.W. (eds.), *Trends in Outdoor Recreation, Leisure and Tourism*. New York: CAB International, pp. 245-255.

Frank, R.E., Massy, W., & Wind, Y. (1972). *Market Segmentation*. Engelwood Cliff: Prentice-Hall.

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

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

Kohonen, T. (1984). *Self Organisation and Associative Memory*. New York: Springer-Verlag.

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

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

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

Silverberg, K.E., Backman, S.J., & Backman, K.F. (1996). A Preliminary Investigation into the Psychographics of Nature-Based Travelers to the Southeastern United States. *Journal of Travel Research*, 35(2), 19-28.

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