



UNIVERSITY  
OF WOLLONGONG  
AUSTRALIA

University of Wollongong  
Research Online

---

Faculty of Commerce - Papers (Archive)

Faculty of Business

---

2003

# Data-Driven Market Segmentation - A Structure-Based Conceptual Framework for Management Decision Support

Sara Dolnicar

*University of Wollongong, s.dolnicar@uq.edu.au*

Friedrich Leisch

*Vienna University of Technology, Austria*

---

## Publication Details

This paper was originally published as: Dolnicar, S & Leisch, F, Data-Driven Market Segmentation - A Structure-Based Conceptual Framework for Managerial Decision Support, Proceedings of the Australia and New Zealand Management Academy Conference, Adelaide, South Australia, 1-3 December 2003.

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: [research-pubs@uow.edu.au](mailto:research-pubs@uow.edu.au)

---

# Data-Driven Market Segmentation - A Structure-Based Conceptual Framework for Management Decision Support

## **Abstract**

Market segmentation increasingly uses homogeneous groups of consumers determined on the basis of empirical market data as target segments (a posteriori-, data-driven-, post hoc segmentation) rather than splitting individuals according to single, typically socio-demographic or geographic, criteria (a priori-, common sense segmentation). A vast amount of contributions has been made to improve methodology of identifying or constructing data-based market segments. However, real world data sets often do not contain clearly separated density clusters. Therefore all techniques used in data-based market segmentation can render multiple solutions of similar quality. So far no attempt has been made to construct a framework enabling managers to systematically choose between different segmentation solutions with regard to their practical usefulness. We propose a framework of such kind.

## **Keywords**

Market segmentation, conceptual framework, stability of data-driven segmentation solutions

## **Disciplines**

Business | Social and Behavioral Sciences

## **Publication Details**

This paper was originally published as: Dolnicar, S & Leisch, F, Data-Driven Market Segmentation - A Structure-Based Conceptual Framework for Managerial Decision Support, Proceedings of the Australia and New Zealand Management Academy Conference, Adelaide, South Australia, 1-3 December 2003.

# Data-Driven Market Segmentation

—

## A Structure-Based Conceptual Framework for Managerial Decision Support.

### Abstract

Market segmentation increasingly uses homogeneous groups of consumers determined on the basis of empirical market data as target segments (*a posteriori*-, data-driven-, *post hoc* segmentation) rather than splitting individuals according to single, typically socio-demographic or geographic, criteria (*a priori*-, common sense segmentation). A vast amount of contributions has been made to improve methodology of identifying or constructing data-based market segments. However, real world data sets often do not contain clearly separated density clusters. Therefore all techniques used in data-based market segmentation can render multiple solutions of similar quality. So far no attempt has been made to construct a framework enabling managers to systematically choose between different segmentation solutions with regard to their practical usefulness. We propose a framework of such kind.

### Keywords

Market segmentation, conceptual framework, stability of data-driven segmentation solutions

Market segmentation is a widely accepted strategic marketing tool in industry. Over time the interest has shifted from *a priori* market segmentation approaches to *a posteriori* procedures (Mazanec 2000) where an empirical data set providing disaggregate consumer information is used to split individuals into homogeneous groups rather than using single variables, typically of socio-demographic or geographic nature. The main focus of development in the area of *a posteriori* segmentation was on increasing methodological sophistication of techniques for identification or construction of such data-driven market segments (Wedel & Kamakura 2002).

However, while methodology improved in sophistication, numerous fundamental questions remain unsolved, leaving managers without guidance with regard to the appropriateness of applying grouping techniques to their data. The core research question underlying this lack of guidance is, which data conditions require which kind of treatment to derive managerially useful market segments.

As Aldenderfer & Blashfield (1984, p.16) put it a long time ago: "The key to using cluster analysis is knowing when these groups are ``real" and not merely imposed on the data by the method."

From the managerial perspective this means that an increased level of transparency is needed, especially when managers have to make the decision, which one of many possible segmentation solutions should be used by the organisation as fundamental strategic building block of the marketing activities.

The aim of this study is to increase transparency by (1) reviewing the implicit concepts underlying major past publications on market segmentation and (2) to propose a framework that guides management through the decision process of which segmentation concept to adopt in a step-wise manner. Using this framework is expected to increase understanding of different data-driven segmentation concepts, to encourage investigation of data structure before choosing a grouping technique and to ease clear articulation of the underlying concept when reporting on a segmentation study to avoid false assumptions about the existence of "true" clusters corresponding to a natural grouping of the data. While this is a conceptual step into providing increased guidance to managers, the next step is to design a toolbox of methods that can be used to investigate data structure and thus fully operationalise the presently conceptual framework.

### **Concepts Implicitly Underlying *A Posteriori* Segmentation Approaches**

While typical journal publications reporting of *a posteriori* segmentations do not explicitly mention whether they have explored the data prior to the grouping task or whether they have any assumptions about the natural existence of groupings or not (Baumann 2000; Dolničar 2002), the major comprehensive books on market segmentation differ in which fundamental assumptions they believe underlie the grouping process, with earlier publications assuming natural groupings and work published in the late 90s and later tending towards acknowledgement of the constructive nature of market segmentation:

Frank, Massy & Wind (1972) state that the purpose of taxonomic procedures is to describe *natural groupings* in empirical data sets. Myers & Tauber (1977) refer to market segments within the field of segmentation research as clearly defined natural groupings of people. Consequently, the goal of the segmentation process is to identify these natural groupings.

Mazanec (1997) does not assume the existence of natural segments, thus implying that homogeneous groups have to be constructed rather than found. Wedel & Kamakura (2002) are in agreement with this latter assumption by saying that market segmentation involves *artificial groupings* of individuals that are constructed for best possible targeting action.

### **A Conceptual Data-Driven Segmentation Framework**

The proposed framework is outlined in Figure 1. It illustrates the three fundamental categories of data-driven segmentation studies, which can be distinguished by asking two sequential questions. The first and most important criterion for contrasting the approaches is to determine whether there is any structure in the empirical data at all or not. Numerous techniques have been proposed in the past to investigate this question (often in the context of determining the correct number of clusters): when clustering approaches are used, scree plots are expected to provide insight into the data structure, dendrogram inspection is used in the case of hierarchical algorithms, random graph tests can be computed or ensemble methods can help to investigate data structure. The latter involve running the cluster algorithm of choice repeatedly on bootstrap samples of the original data set and comparing the obtained solutions. If there exists a number of segments for which we can reproduce the same clustering result on

the bootstrap samples, we conclude that the the data have some sort of cluster structure. The comparison of the different partitions can be done using compliance measures like the Rand Index or ensemble cluster methods like bagged clustering (Leisch 1999).

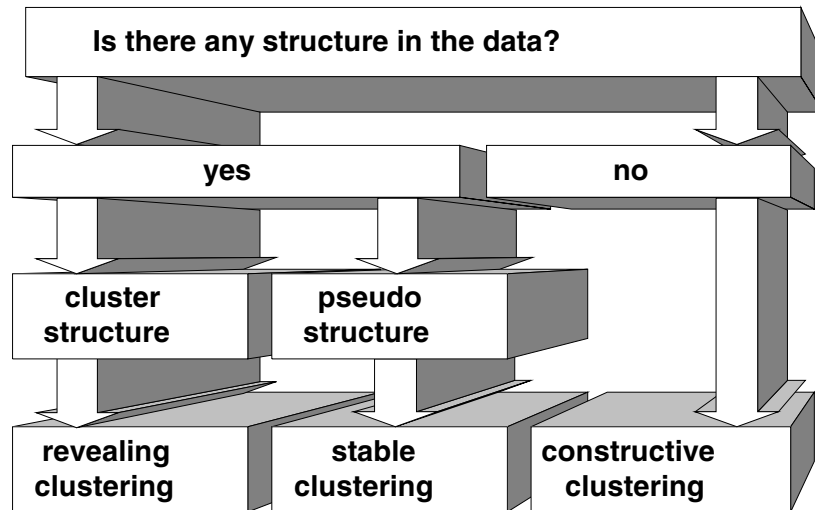


Figure 1: Conceptual Data-Driven Segmentation Framework

If structure is found to exist in the data set, the second question that has to be posed is, of which nature this structure is. Typically, when assuming existence of natural groupings of individuals, it is density cluster structure in multidimensional space that is implied. In this case, and only in this case, the concept of “revealing (or natural) clustering” is appropriate. In line with the assumptions of Frank, Massy & Wind (1972) and Myers & Tauber (1977) natural groupings do exist and the purpose of data-driven market segmentation consequently is to detect and identify these groups. If, however, the structure in the data is not of the nature of density clusters, this goal is an inappropriate one. Other structure in the data may enable the researcher to find stable grouping solutions, where stability means that repeated grouping leads to the same solution(s), we call this pseudo cluster structure. If density clusters do not exist in a otherwise structured data set, “stable clustering” would be aimed at because it clearly is preferable from a managerial perspective to build strategic marketing decisions on stable findings deduced from the data.

The complexity of properly answering this question is illustrated in Figure 2 for the simple case of two-dimensional data sets only. The horizontal axis represents the number of features a mobile phone might have. The vertical axis stands for the price of the phone and the points are respondents’ preferences. Clearly, structure does exist in the data: respondents are not randomly scattered in the two-dimensional space, the data is structured in an ellipse form. However, when solutions searching for three clusters are computed repeatedly, the same cluster centres emerge (located in the area pointed out by circles), misleading the researcher to believe that true clusters have been revealed.

This second question would thus require studying the intensity of criteria used to answer the first question. However, such techniques are not readily available due to a lack of critical values to compare empirical data set criteria levels with. While this second question remains unsatisfactorily operationalised, there are warning signs for the lack of true clusters that

emerge from comprehensive reviews of cluster analysis applications for *a posteriori* segmentation: for instance, structures revealed following the patterns known to be generated by algorithms (Everitt, Landau & Leese 2001) or the number of clusters lying between three and five (Dimitriadou, Dolničar & Weingessel 2002). This phenomenon of cluster numbers lying between three and five in most empirical studies emerged as one of the findings from extensive literature reviews of the use of cluster analysis for the purpose of market segmentation (Baumann, 2000; Dolničar, 2002). In the area of business administration two thirds of all empirical studies claim that three, four or five numbers of clusters are the “correct” or “best” solution. In tourism research two thirds of the investigations result in either three or four clusters. These numbers are found without any systematic association with the nature of the data, the number of variables used, the sample size or any other relevant parameter in the clustering process, which indicates an algorithm-driven occurrence rather than a case-specific data-driven finding. This is the reason that we suggest these cluster numbers to be interpreted with great care and possibly taken as indicator of pseudo- rather than true density cluster structure.

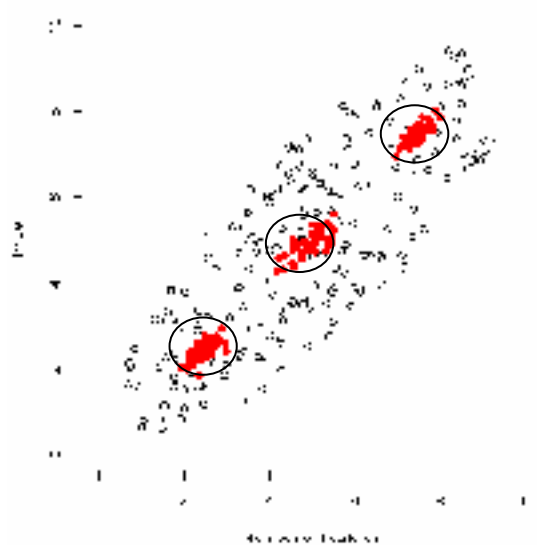


Figure 2: Illustrating the problem of identifying real clusters versus pseudo-structure in a simplified two-dimensional space

If no structure is found in the data set, the only possibility is to choose the “constructive clustering” approach. This basically means that different “random” groupings are compared and the managerially most useful one is chosen. Managerial usefulness can be defined as ease of communication with the segments, ease of identification of these customers in practice etc. It seems that many of the data-driven segmentation studies conducted (see Baumann 2000 for an overview) follow this approach, although they never explicitly state this as it might seem very subjective and unprofessional. However, if there is no structure in the data, this remains the only possibility and management has to be aware of the approach taken and has to decide whether it is better to work with an artificial, constructed grouping of consumers into homogeneous subgroups or whether it is preferable not to group them at all and address the entire market.

## Conclusions and Future Work

In the last decades the popularity of segmentation studies based on empirical data sets has rapidly increased both in academic research as in industry. However, the concepts implicitly

assumed when segmenting consumers have typically not been discussed and transparently revealed.

On the basis of a review of the two basic conceptual foundations of market segmentation found in literature (identification of natural groupings versus construction of artificial groupings) a framework was proposed that systematizes possible approaches on the basis of the structure of the empirical data set. Two questions have to be investigated: (1) Is there any structure in the data and (2) of what nature is that structure. Answering these questions allows users of data-driven segmentation solutions (managers) to determine, which concept of segmentation underlies each particular case. Transparency on the concept allows managers to fully understand the strength of the segmentation solution marketing decisions are based on, where “revealed clustering” represents the most desirable situation and “constructive clustering” is the most arbitrary among the concepts. However, management might choose to prefer a useful arbitrary grouping of consumers into homogeneous segments to a mass market approach as consumer needs will still be more similar to each other and can be met more efficiently.

In any case, it is important to clearly state the segmentation concept underlying each study in order to empower managers to understand what the data basis for their strategic marketing decisions is. Also, the framework is expected to stimulate analysis of data structure before a grouping technique is used.

While this framework represents a conceptual system and a variety of tools to answer the questions posed when proceeding vertically down the tree, future work should focus on providing a comprehensive toolkit of techniques that can be used to most efficiently determine both the existence and nature of the data structure the segmentation analysis is based on.

## References

- Aldenderfer, MS & Blashfield, RK 1984, *Cluster Analysis*, Sage Publications, Beverly Hills.
- Baumann, R 2000, *Marktsegmentierung in den Sozial- und Wirtschaftswissenschaften: eine Metaanalyse der Zielsetzungen und Zugänge*, Vienna University of Economics and Business Administration, Vienna.
- Dimitriadou, E, Dolničar, S & Weingessel, A 2002, ‘An Examination of Indexes for Determining the Number of Clusters in Binary Data Sets’, *Psychometrika*, vol. 67, no. 1, pp. 137-160.
- Dolničar, S 2002, ‘Review of Data-Driven Market Segmentation in Tourism’, *Journal of Travel & Tourism Marketing*, vol. 12, no. 1, pp. 1-22.
- Everitt, BS, Landau, S & Leese, M 2001, *Cluster Analysis*, 4<sup>th</sup> edn, Arnold, London.
- Frank, RE, Massy, WF & Wind, Y 1972, *Market Segmentation*, Prentice-Hall, Englewood Cliffs.
- Haley, RJ 1968, ‘Benefit Segmentation: A Decision-Oriented Research Tool’, *Journal of Marketing*, vol. 32, pp. 30-35.
- Leisch, F 1999, ‘*Bagged Clustering*’, Working Paper No. 51, “Adaptive Information Systems and Modeling in Economics and Management Science”, Vienna University of Economics and Business Administration, Vienna.

Mazanec, JA 1997, 'Segmenting city tourists into vacation styles' in *International City Tourism: Analysis and Strategy*, eds K Grabler, G Maier, J Mazanec & K Wober, Pinter Cassell, London, pp. 114-128).

Mazanec, JA 2000, 'Market Segmentation' in *Encyclopedia of Tourism*, ed J Jafari, Routledge. London, pp. 525.

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

Myers, JH & Tauber, E 1977, *Market structure analysis*, American Marketing Association, Chicago.

Wedel, M & Kamakura, WA 2002, *Market Segmentation. Conceptual and Methodological Foundations.*, Kluwer Academic Publishers, Boston.