Summary of the Book

Market Segmentation Analysis

Understanding It, Doing It,
and Making It Useful

Team Members (click to land on their Github account)

S Deepika(Covered Step-4)

Kunal Aggarwal (Covered Step-5)

Krishnendu Barman (Covered Step-6)

Mohammed Ammaruddin (Covered Step-7)

Gouri Rajendra Uplenchwar(Covered Step-8 & 9)

Step 1:Implications of Committing to Market Segmentation

The key implication is that the needs to commit to the segmentation strategy on the long term. . The commitment to market segmentation goes hand in hand with the willingness and ability of the to make substantial changes (McDonald and Dunbar 1995) and investments.

Modification of existing products, changes in pricing and distribution channels used to sell the product, as well as all communications with the market. These changes, in turn, are likely to influence the internal structure of the , which may need to be adjusted in view of, for example, targeting a handful of different market segments. Croft (1994) recommends that – to maximize the benefits of market segmentation – s need to organize around (p. 66) market, Step 1: Deciding (not) to Segment segments, rather than organizing around products.

3.2 Implementation Barriers:

The first group of barriers relates to senior management. Lack of leadership, pro-active championing, commitment and involvement in the market segmentation process by senior leadership undermines the success of market segmentation. As McDonald and Dunbar (1995, p. 158) state: There can be no doubt that unless the chief executive sees the need for a segmentation review.

Senior management can also prevent market segmentation to be successfully implemented by not making enough resources available, either for the initial market segmentation analysis itself, or for the long-term implementation of a market segmentation strategy. A second group of barriers relates to organisational culture. Lack of market or consumer orientation, resistance to change and new ideas, lack of creative thinking, bad communication and lack of sharing of information and insights across organisational units, short-term thinking,

Another potential problem is lack of training.

Another obstacle may be objective restrictions faced by the , including lack of financial resources, or the inability to make the structural changes required

A company with limited resources needs to pick only the best opportunities to pursue. Process-related barriers include not having clarified the objectives of the market segmentation exercise, lack of planning or bad planning, a lack of structured processes to guide the team through all steps of the market segmentation process, a lack of allocation of responsibilities, and time pressure that stands in the way of trying to find the best possible segmentation outcome.

3 Step 1 Checklist

This first checklist includes not only tasks, but also a series of questions which, if not answered in the affirmative, serve as knock-out criteria. For example: if an is not market-oriented, even the finest of market segmentation analyses cannot be successfully implemented.

Step 2: Specifying the Ideal Target Segment

Abstract

Market segmentation analysis is driven primarily by the desire of an to better cater to a part of the market and, in so doing, secure a competitive advantage. At the end of the segmentation analysis, the needs to select one or more target segments.

After having committed to investigating the value of a segmentation strategy in Step 1, the has to make a major contribution to market segmentation analysis in Step 2. While this contribution is conceptual in nature, it guides many of the following steps, most critically Step 3 (data collection) and Step 8 (selecting one or more target segments). In Step 2 the must determine two sets of segment evaluation criteria. One set of evaluation criteria can be referred to as *knock-out criteria*. These criteria are the essential, non- negotiable features of segments that the would consider targeting. The second set of evaluation criteria can be referred to as *attractiveness criteria*. These criteria are used to evaluate the relative attractiveness of the remaining market segments – those in compliance with the knock-out criteria.

Where knock-out criteria automatically eliminate some of the available market segments, attractiveness criteria are first negotiated by the team, and then applied to determine the overall relative attractiveness of each market segment in Step 8.

2 Knock-Out Criteria

Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria. The first set of such criteria was suggested by Kotler (1994) and includes substantiality, measurability and accessibility (Tynan and Drayton 1987). Kotler himself and a number of other authors have since recommended additional criteria that fall into the knock-out criterion category (Wedel and Kamakura 2000; Lilien and Rangaswamy 2003; McDonald and Dunbar 2012):

- The segment must be **homogeneous**; members of the segment must be similar to one another.
- The segment must be **distinct**; members of the segment must be distinctly different from members of other segments.
- The segment must be **large enough**; the segment must contain enough consumers to make it worthwhile to spend extra money on customizing the marketing mix for them.
- The segment must be **matching** the strengths of the organization; the must have the capability to satisfy segment members' needs.
- Members of the segment must be **identifiable**; it must be possible to spot them in the marketplace.
- The segment must be **reachable**; there has to be a way to get in touch with members of the segment in order to make the customized marketing mix accessible to them.

Knock-out criteria must be understood by senior management, the segmentation team, and the advisory committee. Most of them do not require further specification, but some do. For example, while size is non-negotiable, the exact minimum viable target segment size needs to be specified.

3 Attractiveness Criteria

Attractiveness criteria are not binary in nature. Segments are not assessed as either complying or not complying with attractiveness criteria. Rather, each market segment is rated; it can be more or less attractive with respect to a specific criterion. The attractiveness across all criteria determines whether a market segment is selected as a target segment in Step 8 of market segmentation analysis.

4 Implementing a Structured Process

The most popular structured approach for evaluating market segments in view of selecting them as target markets is the use of a segment evaluation plot.

The segment attractiveness and organisational competitiveness values are determined by the segmentation team. This is necessary because there is no standard set of criteria that could be used by all organizations.

Factors which constitute both segment attractiveness and organisational competitiveness need to be negotiated and agreed upon. To achieve this, a large number of possible criteria has to be investigated before agreement.

There are at least two good reasons to include in this process representatives from a wide range of organisational units. First, each organisational unit has a different perspective on the business of the organisation. As a consequence, members of these units bring different positions to the deliberations.
Secondly, if the segmentation strategy is implemented, it will affect every single unit of the organization. Consequently, all units are key stakeholders of market segmentation analysis.

At the end of this step, the market segmentation team should have a list of approximately six segment attractiveness criteria. Each of these criteria should have a weight attached to it to indicate how important it is to the organization compared to the other criteria.

5 Step 2 Checklist

Task	Who is responsible?	Completed?
Convene a segmentation team meeting.		☐
Discuss and agree on the knock-out criteria of homogeneity, distinctness, size, match, identifiability and reachability. These knock-out criteria will lead to the automatic elimination of market segments which do not comply (in Step 8 at the latest).		☐
Present the knock-out criteria to the advisory committee for discussion and (if required) adjustment.		☐
Individually study available criteria for the assessment of market segment attractiveness.		☐
Discuss the criteria with the other segmentation team members and agree on a subset of no more than six criteria.		☐
Individually distribute 100 points across the segment attractiveness criteria you have agreed upon with the segmentation team. Distribute them in a way that reflects the relative importance of each attractiveness criterion.		☐
Discuss weightings with other segmentation team members and agree on a weighting.		☐
Present the selected segment attractiveness criteria and the proposed weights assigned to each of them to the advisory committee for discussion and (if required) adjustment.		☐

Step 3: Collecting Data

Abstract

The outcome of a market segmentation analysis is only as good as the data upon which it is based.

1 Segmentation Variables

Empirical data forms the basis of both commonsense and data-driven market segmentation. Empirical data is used to identify or create market segments and – later in the process – describe these segments in detail.

segmentation variable to refer to the variable in the empirical data used in commonsense segmentation to split the sample into market segments. In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample. This case is illustrated in Table 5.1. Each row in this table represents one consumer, each variable represents one characteristic of that consumer. An entry of 1 in the data set indicates that the consumer has that characteristic. An entry of 0 indicates that the consumer does not have that characteristic.

All the other personal characteristics available in the data – in this case: age, the number of vacations taken, and information about five benefits people seek or do not seek when they go on vacation – serve as so-called *descriptor variables* . They are used to describe the segments in detail. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio- demographics, but also information about media behavior, allowing marketers to reach their target segment with communication messages.

The difference between commonsense and data-driven market segmentation is that data-driven market segmentation is based not on one, but on multiple segmentation variables. These segmentation variables serve as the starting point for identifying naturally existing, or artificially creating market segments useful to the . Data quality is critical to both (1) assigning each person in the sample to the correct market segment, and (2) being able to correctly describe the segments. The correct description, in turn, makes it possible to develop a customized product, determine the most appropriate pricing strategy, select the best distribution channel, and the most effective communication channel for advertising and promotion.

2 Segmentation Criteria

The term *segmentation criterion* is used here in a broader sense than the term segmentation variable. The term segmentation variable refers to one measured value, for example, one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation. It can also relate to one specific construct, such as benefits sought.

2.3 Psychographic Segmentation

When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, the term psychographic segmentation is used. Most psychographic segmentation studies use a number of segmentation variables, for example: a number of different travel motives, a number of perceived risks when going on vacation.

The psychographic approach has the advantage that it is generally more reflective of the underlying reasons for differences in consumer behavior. For example, tourists whose primary motivation to go on vacation is to learn about other cultures, have a high likelihood of undertaking a cultural holiday at a destination that has ample cultural treasures for them to explore.

2.4 Behavioral Segmentation

A wide range of possible behaviors' can be used for this purpose, including prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion (or across multiple purchase occasions), and information search behavior. In a comparison of different segmentation criteria used as segmentation variables, behaviors reported by tourists emerged as superior to geographic variables.

The key advantage of behavioral approaches is that – if based on actual behavior rather than stated behavior or stated intended behavior – the very behavior of interest is used as the basis of segment extraction.

But behavioral data is not always readily available, especially if the aim is to include in the segmentation analysis potential customers who have not previously purchased the product, rather than limiting oneself to the study of existing customers of the organization.

3 Data from Survey Studies

survey data – as opposed to data obtained from observing actual behavior – can be contaminated by a wide range of biases. Such biases can, in turn, negatively affect the quality of solutions derived from market segmentation analysis. A few key aspects that need to be considered when using survey data are discussed below.

3.1 Choice of Variables

Carefully selecting the variables that are included as segmentation variable in commonsense segmentation, or as segmentation variables in data-driven segmentation, is critical to the quality of the market segmentation solution.

In data-driven segmentation, all variables relevant to the construct captured by the segmentation criterion need to be included. At the same time, unnecessary variables must be avoided.

Unnecessary variables included as segmentation variables divert the attention of the segment extraction algorithm away from information critical to the extraction of optimal market segments.

Redundant items are particularly problematic in the context of market segmentation analysis because they interfere substantially with most segment extraction algorithms' ability to identify correct market segmentation solutions

3.2 Response Options

Answer options provided to respondents in surveys determine the scale of the data available for subsequent analyses. Because many data analytic techniques are based on distance measures , not all survey response options are equally suitable for segmentation analysis.

Options allowing respondents to answer in only one of two ways,

generate *binary* or *dichotomous data* . Such responses can be represented in a data set by 0s and 1s. The distance between 0 and 1 is clearly defined and, as such, poses no difficulties for subsequent segmentation analysis.

3.3 Response Styles

A wide range of response styles manifest in survey answers, including respondents' tendencies to use extreme answer options (STRONGLY AGREE, STRONGLY DISAGREE), to use the midpoint (NEITHER AGREE NOR
DISAGREE), and to agree with all statements. Response styles affect segmentation results because commonly used segment extraction algorithms cannot differentiate between a data entry reflecting the respondent's belief from a data entry reflecting both a respondent's belief and a response style.

3.4 Sample Size

The market segmentation problem in this figure is extremely simple because only two segmentation variables are used. Yet, when the sample size is insufficient (left plot), it is impossible to determine which the correct number of market segments is. If the sample size is sufficient, however (right plot) it is very easy to determine the number and nature of segments in the data set.

The outcome of a market segmentation analysis is only as good as the data upon which it is based. This chapter discusses a range of alternative sources of data that can serve as input for extracting market segments. Key potential dangers associated with each of those sources are discussed. A checklist summarises a number of questions that may assist in ensuring that data of the highest quality is being collected.

1 Segmentation Variables

Empirical data forms the basis of both commonsense and data-driven market segmentation. Empirical data is used to identify or create market segments and – later in the process – describe these segments in detail.

Throughout this book we use the term *segmentation variable* to refer to the variable in the empirical data used in commonsense segmentation to split the sample into market segments. In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample.. Each row in this table represents one consumer, each variable represents one characteristic of that consumer. An entry of 1 in the data set indicates that the consumer has that characteristic. An entry of 0 indicates that the consumer does not have that characteristic. The commonsense segmentation illustrated in Table 5.1 uses gender as the segmentation variable. Market segments are created by simply splitting the sample using this segmentation variable into a segment of women and a segment of men.

Table 5.1 Gender as a possible segmentation variable in commonsense market segmentation

All the other personal characteristics available in the data – in this case: age, the number of vacations taken, and information about five benefits people seek or do not seek when they go on vacation – serve as so-called *descriptor* *variables* . They are used to describe the segments in detail. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. Typical descriptor variables include socio-demographics, but also information about media behavior, allowing marketers to reach their target segment with communication messages.

The difference between commonsense and data-driven market segmentation is that data-driven market segmentation is based not on one, but on multiple segmentation variables. These segmentation variables serve as the starting point for identifying naturally existing, or artificially creating market segments useful to the organization..

Table 5.2 Segmentation variables in data-driven market segmentation

In the data-driven case we may, for example, want to extract market segments of tourists who do not necessarily have gender in common, but rather share a common set of benefits they seek when going on vacation. Sorting the data from Table 5.1 using this set of segmentation variables reveals one segment (shown in the first three rows) characterized by seeking relaxation, culture and meeting people, but not interested in action and exploring. In this case, the benefits sought represent the segmentation variables. The socio- demographic variables, gender, age, and the number of vacations undertaken per annum serve as descriptor variables.

These two simple examples illustrate how critical the quality of empirical data is for developing a valid segmentation solution. When commonsense segments are extracted – even if the nature of the segments is known in advance – data quality is critical to both (1) assigning each person in the sample to the correct market segment, and (2) being able to correctly describe the segments. The correct description, in turn, makes it possible to develop a customized product, determine the most appropriate pricing strategy, select the best distribution channel, and the most effective communication channel for advertising and promotion.

The same holds for data-driven market segmentation where data quality determines the quality of the extracted data-driven market segments, and the quality of the descriptions of the resulting segments. Good market segmentation analysis requires good empirical data .

Empirical data for segmentation studies can come from a range of sources: from survey studies; from observations such as scanner data where purchases are recorded and, frequently, are linked to an individual customer's long-term purchase history via loyalty programs; or from experimental studies.

Optimally, data used in segmentation studies should reflect consumer behavior . Survey data – although it arguably represents the most common source of data for market segmentation studies – can be unreliable in reflecting behavior, especially when the behavior of interest is socially desirable, such as donating money to a charity or behaving in an environmentally friendly way (Karlsson and Dolnicar 2016). Surveys should therefore not be seen as the default source of data for market segmentation studies. Rather, a range of possible sources should be explored. The source that delivers data most closely reflecting actual consumer behavior is preferable.

2 Segmentation Criteria

Long before segments are extracted, and long before data for segment extraction is collected, the organization must make an important decision: it must choose which segmentation criterion to use (Tynan and Drayton 1987). The term *segmentation criterion* is used here in a broader sense than the term segmentation variable. The term segmentation variable refers to one measured value, for example, one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation. It can also relate to one specific construct, such as benefits sought.

The decision which segmentation criterion to use cannot easily be outsourced to either a consultant or a data analyst because it requires prior knowledge about the market. The most common segmentation criteria are geographic, socio-demographic, psychographic and behavioral.

Bock and Uncles (2002) argue that the following differences between consumers are the most relevant in terms of market segmentation: profitability, bargaining power, preferences for benefits or products, barriers to choice and consumer interaction effects. With so many different segmentation criteria available, which is the best to use? As Hoek et al. (1996) note, few guidelines as to the most appropriate base to use in a given marketing context exist (p. 26). Generally, the recommendation is to use the simplest possible approach. Cahill (2006) states this very clearly in his book on lifestyle segmentation (p. 159): Do the least you can. If demographic segmentation will work for your product or service, then use demographic segmentation. If geographic segmentation will work because your product will only appeal to people in a certain region, then use it. Just because psychographic segmentation is sexier and more sophisticated than demographic or geographic segmentation does not make it better. Better is what works for your product or service at the least possible cost.

2.1 Geographic Segmentation

Geographic information is seen as the original segmentation criterion used for the purpose of market segmentation (Lewis et al. 1995; Tynan and Drayton 1987). Typically – when geographic segmentation is used – the consumer's location of residence serves as the only criterion to form market segments. For example: if the national tourism organization of Austria wants to attract tourists from neighboring countries, it needs to use a number of different languages: Italian, German, Slovenian, Hungarian, Czech. Language differences across countries represent a very pragmatic reason for treating tourists from different neighboring countries as different segments.
Interesting examples are also provided by global companies such as Amazon selling its Kindle online: one common web page is used for the description of the base product, then customers are asked to indicate their country of residence and country specific additional information is provided. IKEA offers a similar product range worldwide, yet slight differences in offers, pricing as well as the option to purchase online exist in dependence of the customer's geographic location.

The key advantage of geographic segmentation is that each consumer can easily be assigned to a geographic units.

The key disadvantage is that living in the same country or area does not necessarily mean that people share other characteristics relevant to marketers, such as benefits they seek when purchasing a product. Even in the case of luxury suburbs, it is more likely that socio-demographic criteria are the reason for both similar choice of suburb to live in and similar car preferences. The typical case is best illustrated using tourism: people from the same country of origin are likely to have a wide range of different ideal holidays, depending on whether they are single or travel as a family, whether they are into sports or culture.

Despite the potential shortcomings of using geographic information as the segmentation variable, the location aspect has experienced a revival in international market segmentation studies aiming to extract market segments across geographic boundaries. Such an approach is challenging because the segmentation variable(s) must be meaningful across all the included geographic regions, and because of the known biases that can occur if surveys are completed by respondents from different cultural backgrounds (Steenkamp and Ter Hofstede 2002). An example of such an international market segmentation study is provided by Haverila (2013) who extracted market segments of mobile phone users among young customers across national borders.

2.2 Socio-Demographic Segmentation

Typical socio-demographic segmentation criteria include age, gender, income and education. Socio-demographic segments can be very useful in some industries. For example: luxury goods (associated with high income), cosmetics (associated with gender; even in times where men are targeted, the female and male segments are treated distinctly differently), baby products (associated with gender), retirement villages (associated with age), tourism resort products (associated with having small children or not).

As is the case with geographic segmentation, socio-demographic segmentation criteria have the advantage that segment membership can easily be determined for every consumer. In some instances, the socio-demographic criterion may also offer an explanation for specific product preferences (having children, for example, is the actual reason that families choose a family vacation village where previously, as a couple, their vacation choice may have been entirely different). But in many instances, the socio- demographic criterion is not the *cause* for product preferences, thus not providing sufficient market insight for optimal segmentation decisions. Haley (1985) estimates that demographics explain about 5% of the variance in consumer behavior. Yankelovich and Meer (2006) argue that socio- demographics do not represent a strong basis for market segmentation, suggesting that values, tastes and preferences are more useful because they are more influential in terms of consumers' buying decisions.

2.3 Psychographic Segmentation

When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, the term psychographic segmentation is used. Haley (1985) explains that the word psychographics was intended as an umbrella term to cover all measures of the mind (p. 7). Benefit segmentation, which Haley (1968) is credited for, is arguably the most popular kind of psychographic segmentation. Lifestyle segmentation is another popular psychographic segmentation approach (Cahill 2006); it is based on people's activities, opinions and interests.

Psychographic criteria are, by nature, more complex than geographic or socio- demographic criteria because it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest. As a consequence, most psychographic segmentation studies use a number of segmentation variables, for example: a number of different travel motives, a number of perceived risks when going on vacation.

The psychographic approach has the advantage that it is generally more reflective of the underlying reasons for differences in consumer behavior. For example, tourists whose primary motivation to go on vacation is to learn about other cultures, have a high likelihood of undertaking a cultural holiday at a destination that has ample cultural treasures for them to explore. Not surprisingly, therefore, travel motives have been frequently used as the basis for data-driven market segmentation in tourism (Bieger and Laesser 2002; Laesser et al. 2006; Boksberger and Laesser 2009). The disadvantage of the psychographic approach is the increased complexity of determining segment memberships for consumers. Also, the power of the psychographic approach depends heavily on the reliability and validity of the empirical measures used to capture the psychographic dimensions of interest.

2.4 Behavioral Segmentation

Another approach to segment extraction is to search directly for similarities in behavior or reported behavior. A wide range of possible behaviors can be used for this purpose, including prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion (or across multiple purchase occasions), and information search behavior. In a comparison of different segmentation criteria used as segmentation variables, behaviors reported by tourists emerged as superior to geographic variables (Moscardo et al. 2001).

The key advantage of behavioral approaches is that – if based on actual behavior rather than stated behavior or stated intended behavior – the very behavior of interest is used as the basis of segment extraction. As such, behavioral segmentation groups people by the similarity which matters most. Examples of such segmentation analyses are provided by Tsai and Chiu (2004) who use actual expenses of consumers as segmentation variables, and Heilman and Bowman (2002) who use actual purchase data across product categories. Brand choice behavior over time has also been used as segmentation variable by several authors (Poulsen 1990; Bockenholt and Langeheine 1996; Ramaswamy 1997, see also Section 7.3.3). Using behavioral data also avoids the need for the development of valid measures for psychological constructs.

But behavioral data is not always readily available, especially if the aim is to include in the segmentation analysis potential customers who have not previously purchased the product, rather than limiting oneself to the study of existing customers of the organization.

3 Data from Survey Studies

Most market segmentation analyses are based on survey data. Survey data is cheap and easy to collect, making it a feasible approach for any organization. But survey data – as opposed to data obtained from observing actual behavior
– can be contaminated by a wide range of biases. Such biases can, in turn, negatively affect the quality of solutions derived from market segmentation analysis. A few key aspects that need to be considered when using survey data are discussed below.

3.1 Choice of Variables

Carefully selecting the variables that are included as segmentation variable in commonsense segmentation, or as segmentation variables in data-driven segmentation, is critical to the quality of the market segmentation solution.

In data-driven segmentation, all variables relevant to the construct captured by the segmentation criterion need to be included. At the same time, unnecessary variables must be avoided. Including unnecessary variables can make questionnaires long and tedious for respondents, which, in turn, causes respondent fatigue . Fatigued respondents tend to provide responses of lower quality (Johnson et al. 1990; Dolnicar and Rossiter 2008). Including unnecessary variables also increases the dimensionality of the segmentation problem without adding relevant information, making the task of extracting market segments unnecessarily difficult for any data analytic technique. The issue of the appropriate ratio of the number of variables and the available sample is discussed later in this chapter. Unnecessary variables included as segmentation variables divert the attention of the segment extraction algorithm away from information critical to the extraction of optimal market segments. Such variables are referred to as *noisy variables* or *masking variables* and have been repeatedly shown to prevent algorithms from identifying the correct segmentation solution (Brusco 2004; Carmone et al. 1999; DeSarbo et al. 1984; DeSarbo and Mahajan 1984; Milligan 1980).

Noisy variables do not contribute any information necessary for the identification of the correct market segments. Instead, their presence makes it more difficult for the algorithm to extract the correct solution. Noisy variables can result from not carefully developing survey questions, or from not carefully selecting segmentation variables from among the available survey items. The problem of noisy variables negatively affecting the segmentation solution can be avoided at the data collection and the variable selection stage of market segmentation analysis.

The recommendation is to ask all necessary and unique questions, while resisting the temptation to include unnecessary or redundant questions. Redundant questions are common in survey research when scale development follows traditional psychometric principles (Nunally 1978), as introduced to marketing most prominently by Churchill (1979). More recently, Rossiter (2002, 2011) has questioned this practice, especially in the context of measuring concrete objects and attributes that are interpreted consistently as meaning the same by respondents. Redundant items are particularly problematic in the context of market segmentation analysis because they interfere substantially with most segment extraction algorithms' ability to identify correct market segmentation solutions (Dolnicar et al. 2016).

Developing a good questionnaire typically requires conducting exploratory or qualitative research. Exploratory research offers insights about people's beliefs that survey research cannot offer. These insights can then be categorized and included in a questionnaire as a list of answer options. Such a two-stage process involving both qualitative, exploratory and quantitative survey research ensures that no critically important variables are omitted.

3.2 Response Options

Answer options provided to respondents in surveys determine the scale of the data available for subsequent analyses. Because many data analytic techniques are based on distance measures , not all survey response options are equally suitable for segmentation analysis.

Options allowing respondents to answer in only one of two ways,

generate *binary* or *dichotomous data* . Such responses can be represented in a data set by 0s and 1s. The distance between 0 and 1 is clearly defined and, as such, poses no difficulties for subsequent segmentation analysis. Options allowing respondents to select an answer from a range of unordered categories correspond to *nominal variables* . If asked about their occupation, repondents can select only one option from a list of unordered options.
Nominal variables can be transformed into binary data by introducing a binary variable for each of the answer options.

Options allowing respondents to indicate a number, such as age or nights stayed at a hotel, generate *metric data* . Metric data allow any statistical procedure to be performed (including the measurement of distance), and are therefore well suited for segmentation analysis. The most commonly used response option in survey research, however, is a limited number of ordered answer options larger than two. Respondents are asked, for example, to express – using five or seven response options – their agreement with a series of statements. This answer format generates *ordinal data* , meaning that the options are ordered. But the distance between adjacent answer options is not clearly defined. As a consequence, it is not possible to apply standard distance measures to such data, unless strong assumptions are made. Step 5 provides a detailed discussion of suitable distance measures for each scale level.

Preferably, therefore, either metric or binary response options should be provided to respondents if those options are meaningful with respect to the question asked. Using binary or metric response options prevents subsequent complications relating to the distance measure in the process of data-driven segmentation analysis. The visual analogue scale allows respondents to indicate a position along a continuous line between two end-points, and leads to data that can be assumed to be metric. The visual analogue scale has experienced a revival with the popularity of online survey research, where it is frequently used and referred to as a slider scale . In many contexts, binary response options have been shown to outperform ordinal answer options (Dolnicar 2003; Dolnicar et al. 2011, 2012).

3.3 Response Styles

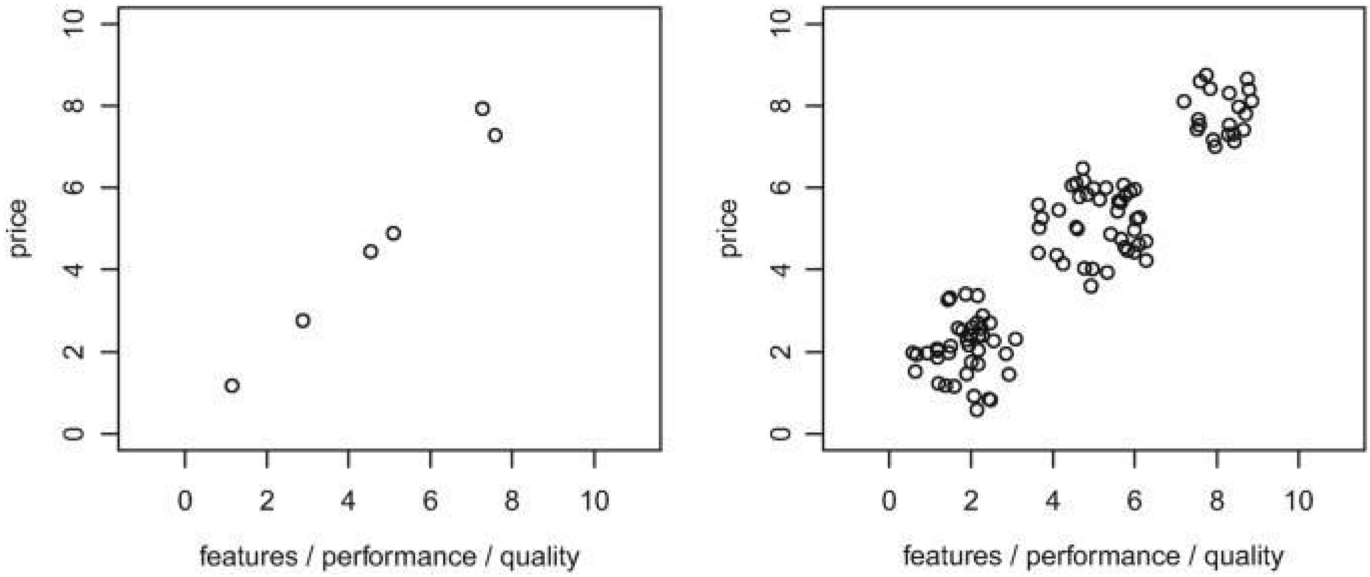
Survey data is prone to capturing biases. A response bias is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content (i.e., what the items were designed to measure) (Paulhus 1991, p. 17). If a bias is displayed by a respondent consistently over time, and independently of the survey questions asked, it represents a response style .

A wide range of response styles manifest in survey answers, including respondents' tendencies to use extreme answer options (STRONGLY AGREE, STRONGLY DISAGREE), to use the midpoint (NEITHER AGREE NOR DISAGREE), and to agree with all statements. Response styles affect segmentation results because commonly used segment extraction algorithms cannot differentiate between a data entry reflecting the respondent's belief from a data entry reflecting both a respondent's belief and a response style. For example, some respondents displaying an acquiescence bias (a tendency to agree with all questions) could result in one market segment having much higher than average agreement with all answers. Such a segment could be misinterpreted. Imagine a market segmentation based on responses to a series of questions asking tourists to indicate whether or not they spent money on certain aspects of their vacation, including DINING OUT, VISITING THEME PARKS, USING PUBLIC TRANSPORT, etc. A market segment saying YES to all those items would, no doubt, appear to be highly attractive for a tourist destination holding the promise of the existence of a high-spending tourist segment. It could equally well just reflect a response style. It is critical, therefore, to minimise the risk of capturing response styles when data is collected for the purpose of market segmentation. In cases where attractive market segments emerge with response patterns potentially caused by a response style, additional analyses are required to exclude this possibility. Alternatively, respondents affected by such a response style must be removed before choosing to target such a market segment.

3.4 Sample Size

Many statistical analyses are accompanied by sample size recommendations. Not so market segmentation analysis. Figure 5.1 illustrates the problem any segmentation algorithm faces if the sample is insufficient. The market segmentation problem in this figure is extremely simple because only two segmentation variables are used. Yet, when the sample size is insufficient (left plot), it is impossible to determine which the correct number of market segments is. If the sample size is sufficient, however (right plot) it is very easy to determine the number and nature of segments in the data set.

Fig. 5.1



Illustrating the importance of sufficient sample size in market segmentation analysis

4 Data from Internal Sources

Increasingly organizations have access to substantial amounts of internal data that can be harvested for the purpose of market segmentation analysis. Typical examples are available to grocery stores, booking data available through airline loyalty programs, and online purchase data advantage is that such data are usually automatically generated and – if organizations are capable of storing data in a format that makes them easy to access – no extra effort is required to collect data.

The danger of using internal data is that it may be systematically biased by over-representing existing customers.

5 Data from Experimental Studies

Experimental data can result from field or laboratory experiments. For example, they can be the result of tests how people respond to certain advertisements. The response to the advertisement could then be used as a segmentation criterion. Experimental data can also result from choice experiments or conjoint analyses .

Step 4: Exploring Data

1 A First Glimpse at the Data

After data collection, exploratory data analysis cleans and – if necessary – pre- processes the data. This exploration stage also offers guidance on the most suitable algorithm for extracting meaningful market segments.

At a more technical level, data exploration helps to (1) identify the measurement levels of the variables; (2) investigate the univariate distributions of each of the variables; and (3) assess dependency structures between variables.

2 Data Cleaning

The first step before commencing data analysis is to clean the data. This includes checking if all values have been recorded correctly, and if consistent labels for the levels of categorical variables have been used.

Returning to the Australian travel motives data set, the summary for the variables Gender and Age indicates that no data cleaning is required for these variables. The summary of the variable Income2 reveals that the categories are not sorted in order. This is a consequence of how data is read into R. R functions like read.csv() or read.table() convert columns containing information other than numbers into factors. Factors are the default format for storing categorical variables in R.

3 Descriptive Analysis

Descriptive numeric and graphic representations provide insights into the data. Statistical software packages offer a wide variety of tools for descriptive analysis. In R, we obtain a numeric summary of the data with command `summary()`. This command returns the range, the quartiles, and the mean for numeric variables.

Histograms visualize the distribution of numeric variables. They show how often observations within a certain value range occur. Histograms reveal if the distribution of a variable is unimodal and symmetric or skewed.

4 Pre-Processing

4.1 Categorical Variables

Two pre-processing procedures are often used for categorical variables. One is merging levels of categorical variables before further analysis, the other one is converting categorical variables to numeric ones, if it makes sense to do so.

Merging levels of categorical variables is useful if the original categories are too differentiated (too many). Thinking back to the income variables, for example, the original income variable as used in the survey.

4.2 Numeric Variables

The range of values of a segmentation variable affects its relative influence in distance-based methods of segment extraction. If, for example, one of the segmentation variables is binary (with values 0 or 1 indicating whether or not a tourist likes to dine out during their vacation), and a second variable indicates the expenditure in dollars per person per day (and ranges from zero to $1000), a difference in spend per person per day of one dollar is weighted equally as the difference between liking to dine out or not. To balance the influence of segmentation variables on segmentation results, variables can be standardized. Standardizing variables means transforming them in a way that puts them on a common scale.

The default standardization method in statistics subtracts the empirical mean $\bar{x}$ and divides by the empirical standard deviation s.

5 Principal Components Analysis

Principal components analysis (PCA) transforms a multivariate data set containing metric variables to a new data set with variables – referred to as principal components – which are uncorrelated and ordered by importance. The first variable (principle component) contains most of the variability, the second principle component contains the second most variability, and so on.

Principal components analysis works off the covariance or correlation matrix of several numeric variables. If all variables are measured on the same scale, and have similar data ranges, it is not important which one to use.

6 Step 4 Checklist

Task	Who is responsible?	Completed?
Explore the data to determine if there are any inconsistencies and if there are any systematic contaminations.		☐
If necessary, clean the data.		☐
If necessary, pre-process the data.		☐
Check if the number of segmentation variables is too high given the available sample size. You should have information from a minimum of 100 consumers for each segmentation variable.		☐
If you have too many segmentation variables, use one of the available approaches to select a subset.		☐
Check if the segmentation variables are correlated. If they are, choose a subset of uncorrelated segmentation variables.		☐
Pass on the cleaned and pre-processed data to Step 5 where segments will be extracted from it.		☐

Implications of committing to Market Segmentation

Market segmentation can be a key marketing strategy, but it requires a long-term commitment and substantial investments in research, product development, pricing, distribution channels, and communication. To maximize the benefits of market segmentation, organizations need to organize around market segments, rather than products. The decision to pursue a market segmentation strategy should be made at the highest execution level and systematically communicated and reinforced across all organizational levels and units.

Implementation Barriers

Several barriers to the successful implementation of a market segmentation strategy in organizations are identified in various books, including lack of leadership and resources from senior management, resistance to change and lack of market orientation in the organizational culture, lack of training and expertise, objective restrictions, process-related barriers, and difficulty in understanding and interrupting results. These barriers can be proactively removed, but if not, abandoning the attempt should be considered. To succeed, dedication, patience, and a willingness to appreciate the challenges are required.

Specifying the Ideal Target Segment

Segment Evaluation Criteria

The third layer of market segmentation analysis relies heavily on user input throughout the process rather than just at the beginning or end. The organization must determine two sets of criteria for segment evaluation: knock-out criteria (essential, non-negotiable features) and attractiveness criteria (used to evaluate the relative attractiveness of remaining segments). The literature proposes a wide array of possible criteria, but the segmentation team must select which ones to use and assess their relative importance. Knock-out criteria automatically eliminate some segments, while attractiveness criteria are negotiated and then applied to determine the overall relative attractiveness of each market segment.

Knock-Out Criteria

Knock-out criteria are used to determine if market segments qualify for assessment using segment attractiveness criteria. The criteria include homogeneity, distinctiveness, size, matching strengths of the organization, identifiability, and reachability. These criteria are non-negotiable and must be understood by senior management, the segmentation team, and the advisory committee. The exact minimum viable target segment size needs to be specified.

Implementing a Structured Process

The article discusses the importance of a structured approach to evaluating market segments and recommends the use of a segment evaluation plot to assess segment attractiveness and organizational competitiveness. It suggests that a team of people should determine the criteria for both these factors, and representatives from various organizational units should be included in the process. The article emphasizes the importance of selecting attractiveness criteria at an early stage in the process to facilitate data collection and make target segment selection easier. It also recommends allocating weights to each criterion based on their relative importance, through negotiation and agreement among team members and approval from the advisory committee.

Collecting Data

Segmentation Variables

Table 5.1 Gender as a possible segmentation variable in commonsense market segmentation

Sociodemographics		Travel behaviour	Benefits sought				
gender	age	Nº of vacations	relaxation	action	culture	explore	meet people
Female	34	2	1	0	1	0	1
Female	55	3	1	0	1	0	1
Female	68	1	0	1	1	0	0
Female	34	1	0	0	1	0	0
Female	22	0	1	0	1	1	1
Female	31	3	1	0	1	1	1
Male	87	2	1	0	1	0	1
Male	55	4	0	1	0	1	1
Male	43	0	0	1	0	1	0
Male	23	0	0	1	1	0	1
Male	19	3	0	1	1	0	1
Male	64	4	0	0	0	0	0
segmentation variable	descriptor variables						

Table 5.2 Segmentation variables in data-driven market segmentation

Sociodemographics		Travel behaviour	Benefits sought				
gender	age	Nº of vacations	relaxation	action	culture	explore	meet people
Female	34	2	1	0	1	0	1
Female	55	3	1	0	1	0	1
Male	87	2	1	0	1	0	1
Female	68	1	0	1	1	0	0
Female	34	1	0	0	1	0	0
Female	22	0	1	0	1	1	1
Female	31	3	1	0	1	1	1
Male	55	4	0	1	0	1	1
Male	43	0	0	1	0	1	0
Male	23	0	0	1	1	0	1
Male	19	3	0	1	1	0	1
Male	64	4	0	0	0	0	0
descriptor variables			segmentation variables				

The article discusses the importance of empirical data in market segmentation, which forms the basis of both commonsense and data-driven approaches. In commonsense segmentation, a single characteristic is used as the segmentation variable to split the sample into market segments. In contrast, data-driven segmentation involves multiple segmentation variables to identify naturally existing or artificially created market segments. Good empirical data is critical for developing a valid segmentation solution, and it can come from a range of sources, such as survey studies, observations, and experimental studies. The source that delivers data most closely reflecting actual consumer behavior is preferable. The quality of empirical data determines the quality of the extracted market segments and the quality of the descriptions of the resulting segments, which is critical for developing a customised product, pricing strategy, distribution channel, and communication channel for advertising and promotion.

Segmentation Variables

Before data for market segmentation is collected, an organization must decide which segmentation criterion to use. The most common criteria are geographic, socio-demographic, psychographic, and behavioral. There are many different segmentation criteria available, but the recommendation is to use the simplest possible approach that works for the product or service at the least possible cost. The decision cannot be outsourced and requires prior knowledge about the market. The relevant differences between consumers for market segmentation are profitability, bargaining power, preferences, barriers to choice, and interaction effects.

Geographic Segmentation

Geographic segmentation is the original and simplest segmentation criterion used in market segmentation, where a consumer's location of residence serves as the only criterion to form market segments. It is useful when targeting consumers in specific regions or countries, but it may not account for other important characteristics relevant to marketers. Despite its potential shortcomings, geographic information has experienced a revival in international market segmentation studies aiming to extract market segments across geographic boundaries.

Socio-Demographic Segmentation

Socio-demographic segmentation criteria, such as age, gender, income, and education, can be useful in certain industries, but they may not always provide sufficient insight into consumer behavior and preferences. While demographic factors may explain some variance in consumer behavior, they are not always the primary cause of product preferences. Yankelovich and Meer (2006) argue that values, tastes, and preferences are more influential in consumers' buying decisions and, therefore, may be a more useful basis for market segmentation. Haley (1985) estimates that demographics only explain about 5% of the variance in consumer behavior.

Psychographic Segmentation

Psychographic segmentation is a grouping of people according to psychological criteria such as beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. It is a more complex approach than geographic or socio-demographic criteria because it is difficult to find a single characteristic that will provide insight into the psychographic dimension of interest. Benefit and lifestyle segmentation are popular psychographic segmentation approaches. The psychographic approach has the advantage of reflecting the underlying reasons for differences in consumer behavior. However, determining segment memberships for consumers can be complex, and the reliability and validity of the empirical measures used to capture psychographic dimensions are crucial.

Behavioural Segmentation

Behavioural segmentation is an approach to segment extraction that groups people based on their actual behaviour or reported behaviour, such as prior experience with the product, frequency of purchase, and information search behaviour. This approach is advantageous because it uses the very behaviour of interest as the basis for segmentation, without the need for the development of valid measures for psychological constructs. However, behavioural data may not always be readily available, especially when targeting potential customers who have not previously purchased the product.

Choice of Variables

In both commonsense and data-driven segmentation, it is crucial to carefully select the variables included as segmentation variables to ensure the quality of the segmentation solution. In data-driven segmentation, all relevant variables must be included while avoiding unnecessary variables that can cause respondent fatigue and make the extraction of optimal segments more difficult. Noisy variables or masking variables, which do not contribute to identifying the correct market segments, can negatively affect the segmentation solution. To avoid these issues, it is recommended to ask all necessary and unique questions while avoiding redundant questions. A good questionnaire requires conducting exploratory or qualitative research to ensure that no critical variables are omitted.

Response Options

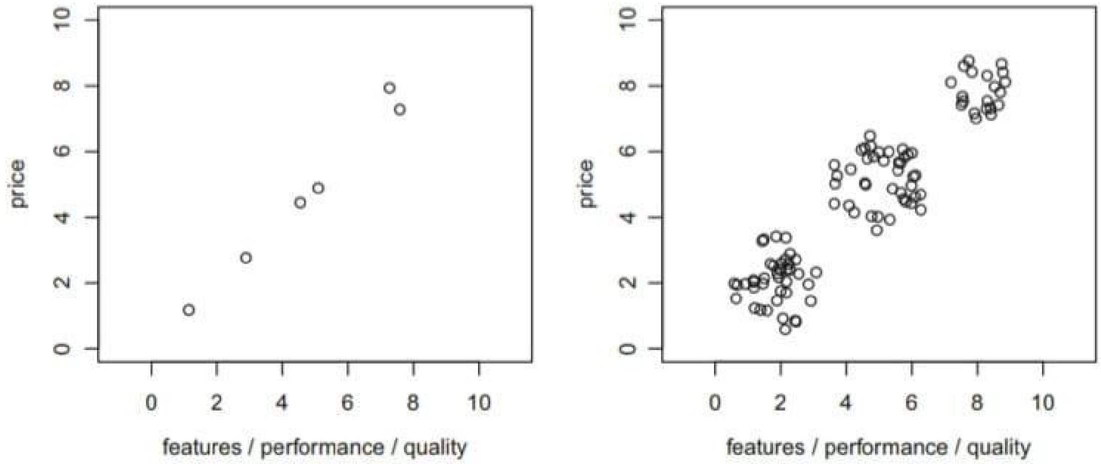
The response options provided in surveys can affect the suitability of data for segmentation analysis. Binary and metric data are preferred, as they allow for clear distance measures. Nominal and ordinal data are less suitable for segmentation analysis, as the distance between answer options is not clearly defined. Visual analogue scales can be used to capture fine nuances of responses. Binary response options have been shown to outperform ordinal options, especially when formulated in a level-free way.

Response Styles

Survey data is vulnerable to response biases, which can result in response styles that affect segmentation analysis. Response biases can cause respondents to answer questions in a consistent way that is not related to the specific content of the questions. These biases can lead to misleading segmentation results, making it essential to minimize the risk of capturing response styles when collecting data for market segmentation. Identifying and removing respondents affected by response styles can help prevent this problem. Additional analyses may also be necessary to exclude the possibility of response style bias in attractive market segments.

Sample Size

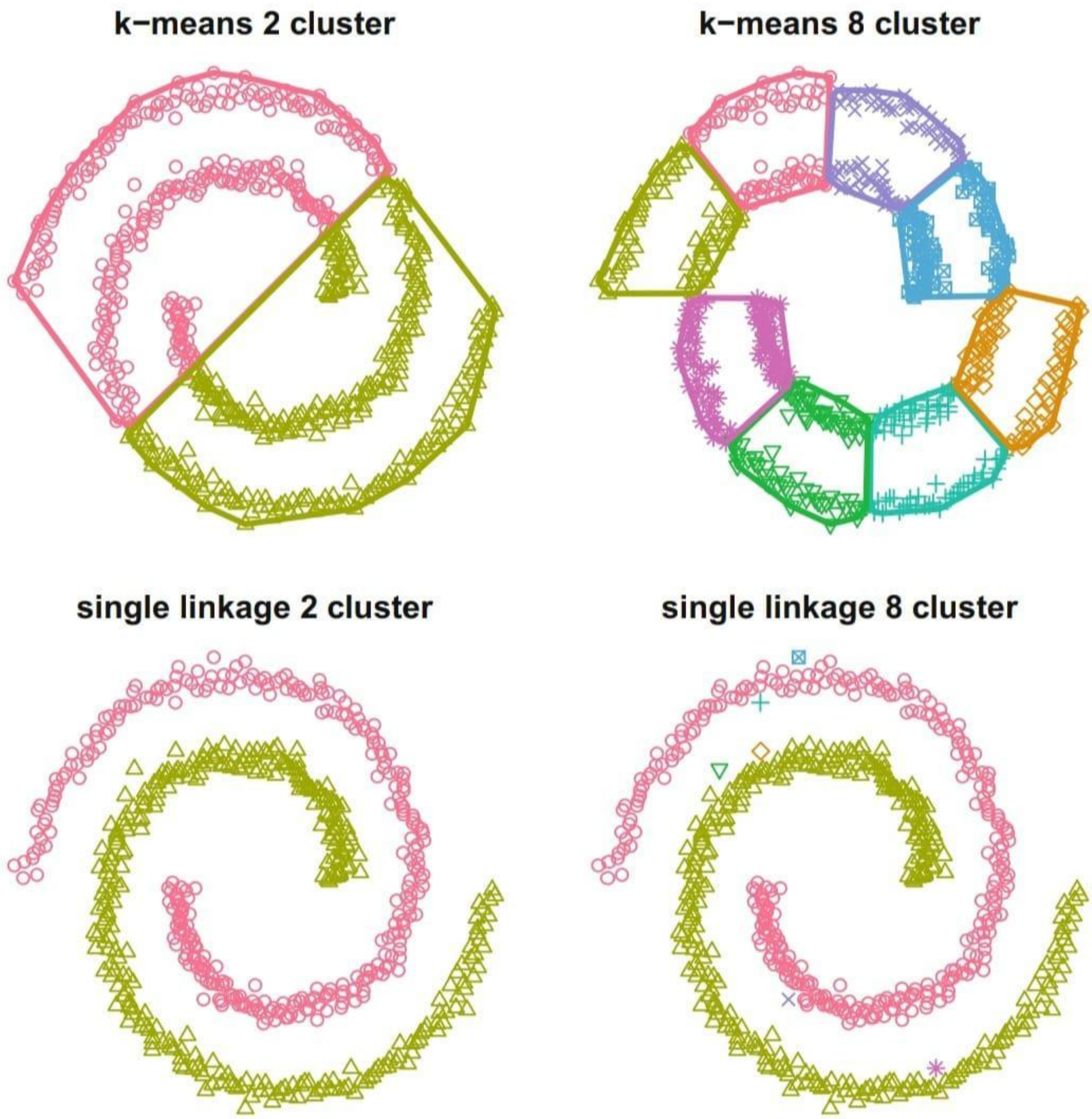
The issue of sample size in market segmentation analysis is crucial. Insufficient sample size can make it impossible to determine the correct number of market segments, while sufficient sample size makes it easy to identify both the number and nature of segments. Different studies have proposed different rules of thumb for sample size, ranging from at least $2p$ (or five times $2p$) to $10 \cdot p \cdot k$, where p is the number of segmentation variables and k is the number of segments.



Simulation studies have shown that sample size has a significant effect on the correctness of segment recovery, as measured by the adjusted Rand index, which assesses the congruence between two segmentation solutions. Overall, market segmentation analysis requires careful consideration of sample size to ensure the accuracy and validity of the results.

Extracting Segments

The article discusses the exploratory nature of data-driven market segmentation analysis, which is often based on unstructured consumer data. The choice of segmentation method strongly affects the resulting segmentation solution, as algorithms impose structure on the extracted segments. Cluster analysis is a common method used, and selecting a suitable clustering method requires matching the data analytic features with context-dependent requirements. The article provides an illustrative example of how different algorithms impose structure on the extracted segments, using k-means cluster analysis and a dataset containing two spiralling segments.



The article continues to discuss the example in Fig. 7.1, comparing the segmentation solutions obtained from k-means and single linkage hierarchical clustering algorithms. Single linkage correctly identifies the two spiralling segments in the data, while k-means fails to do so because it is designed to construct compact, round clusters. However, the article notes that there is no single best algorithm for all data sets, and the choice of algorithm depends on the structure and separation of the data. If data is not well-structured, the tendency of the algorithm will strongly influence the segmentation solution.

This chapter provides an overview of the most popular methods used in market segmentation, including distance-based and model-based methods. There is no single best algorithm for all situations, and each method has advantages and disadvantages. It is important to investigate and compare alternative segmentation solutions to arrive at a good final solution, considering data characteristics and expected or desired segment characteristics. Table 7.1 provides information to guide algorithm selection, such as the size of the available data set, the expected number and size of segments, and the scale level of the segmentation variables. Other special structures of the data can also restrict the set of suitable algorithms.

Distance Measures

Table 7.2 Artificial data set on tourist activities: percentage of time spent on three activities

	beach	action	culture
Anna	100	0	0
Bill	100	0	0
Frank	60	40	0
Julia	70	0	30
Maria	80	0	20
Michael	0	90	10
Tom	50	20	30

The given text explains the concept of measuring the distance between two vectors (representing observations) in market segmentation analysis. The distance is a function with two arguments - two vectors x and y - and the result is the distance between them. The text explains three common distance measures used in market segmentation analysis: Euclidean distance, Manhattan or absolute distance, and asymmetric binary distance. Euclidean distance measures the direct straight-line distance between two points in two-dimensional space, while Manhattan distance gives the distance assuming streets on a grid (like in Manhattan) need to be used to get from one point to another. Both Euclidean and Manhattan distances use all dimensions of the vectors x and y. The text also mentions the criteria that a distance measure must comply with, such as symmetry, the distance of a vector to itself being 0, and the triangle inequality.

d(x, y)=d(y, x)   Symmetric Criteria

d(x, y)=0

Therefore, x=y

In addition, most distance measures fulfil the so-called triangle inequality:

$$D(x, z) < d(x, y) + d(y, z)$$

The triangle inequality says that if one goes from x to z with an intermediate stop in y, the combined distance is at least as long as going from x to z directly.

Hierarchical Methods

Hierarchical clustering methods are commonly used to group data in market segmentation analysis. These methods involve grouping observations or consumers into clusters, starting from either the complete data set or with each observation representing its own cluster. The process continues by either splitting or merging clusters until the complete data set forms one large market segment or until each observation represents its own market segment.

Agglomerative hierarchical clustering and divisive hierarchical clustering are two approaches used to group data. In agglomerative hierarchical clustering, each observation initially represents its own cluster, and then the two closest clusters are merged. This process continues until all the clusters form one large segment. Conversely, in divisive hierarchical clustering, the complete data set starts as one big market segment, which is split into two market segments in the first step. The two segments are then split into two new segments, and this process continues until each consumer has its own segment.

Single linkage: distance between the two closest observations of the two sets.

$l(X, Y) = \min_{x \in X, y \in Y} d(x, y)$

Complete linkage: distance between the two observations of the two sets that are farthest away from each other.

$l(X, Y) = \max_{x \in X, y \in Y} d(x, y)$

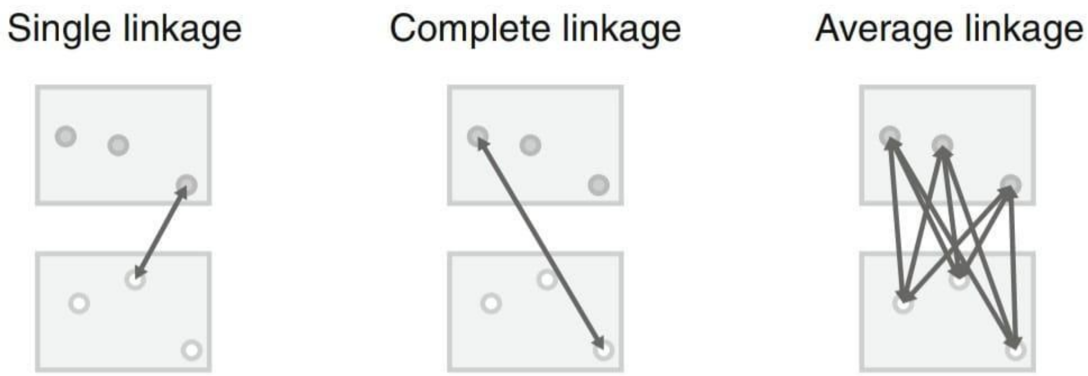
Average linkage: mean distance between observations of the two sets.

$l(X, Y) = 1/|X||Y| \sum_{x \in X}, \sum_{y \in Y} d(x, y)$.

The different approaches result in a sequence of nested partitions ranging from partitions containing only one group to n groups. In general, hierarchical clustering is an exploratory technique used to reveal different features of the data. Different combinations of distance measures and linkage methods can be used in clustering to reveal different features of the data.

Several linkage methods are available, including single linkage, complete linkage, and average linkage. Single linkage uses a “next neighbor” approach to join sets, meaning that the two closest consumers are united. This method is useful for revealing non-convex, non-linear structures in data. Complete and average linkage are better at extracting more compact clusters.

Another popular clustering method is Ward clustering, which is based on squared Euclidean distances. It involves joining sets of observations based on the minimal weighted squared Euclidean distance between cluster centers. The resulting clusters' midpoints are the segment representatives. The hierarchical clustering method's output is presented as a dendogram, which is a tree diagram representing the hierarchy of market segments formed at each step of the procedure.



A comparison of different linkage methods between two sets of points

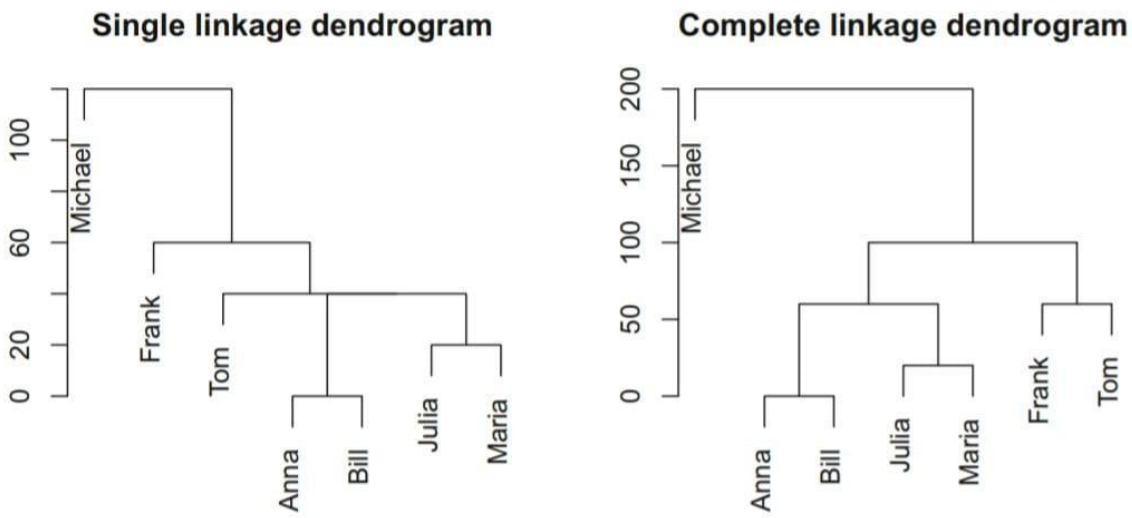


Fig. 7.4 Single and complete linkage clustering of the tourist data shown in Table 7.2

Partitioning Methods

Hierarchical clustering is effective for small data sets with up to a few hundred observations, but becomes impractical for larger data sets due to the difficulty of reading dendrograms and the memory requirements for pairwise distance matrices. For larger data sets, clustering methods that create a single partition are more suitable, as they only need to calculate distances between each observation and the centers of segments. Partitioning clustering algorithms can extract a specific number of segments with fewer distance calculations than hierarchical clustering, making them more efficient for larger data sets. It is better to optimize specifically for the goal of extracting a few segments rather than building a complete dendogram and cutting it into segments heuristically.

K- Means and K-Centroid Clustering

The k-means clustering method is a popular partitioning method that divides a set of observations (consumers) into subsets (market segments) based on their similarity. The algorithm uses a representative value called the centroid, which is the column-wise mean of all members in the segment. The goal is to group consumers into a given number of segments so that they are similar to their fellow segment members, but dissimilar to members of other segments. The algorithm is iterative, improving the partition in each step and bound to converge, but not necessarily to the global optimum. R function kmeans() implements different algorithms for k-means clustering using the squared Euclidean distance, and flexclust provides a generalization to other distance measures.

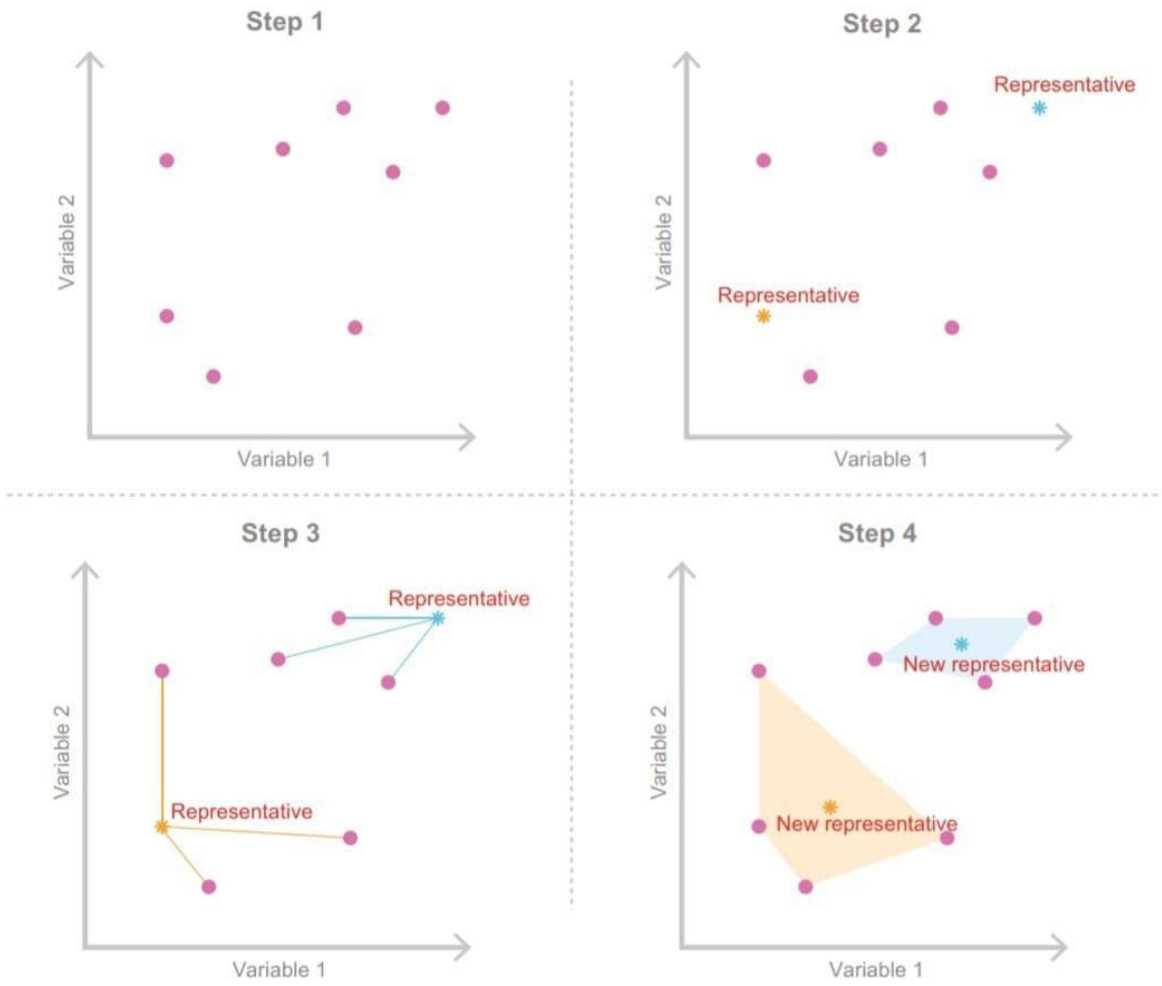


Fig. 7.7 Simplified visualisation of the *k*-means clustering algorithm

"Improved" K-Means

To improve the k-means clustering algorithm, it is recommended to use "smart" starting values instead of randomly drawing k consumers from the data set. Randomly drawn consumers may not be representative of the data space and can increase the likelihood of the algorithm getting stuck in a local optimum. One way to avoid this is to initialise the algorithm using starting points that are evenly spread across the entire data space. Steinley and Brusco (2007) compared 12 different strategies for initialising the k-means algorithm and found that the best approach is to randomly draw many starting points and select the best set based on their ability to represent the data. Good representatives are those that are close to their segment members, while bad representatives are far away from their segment members.

Hard Competitive Learning

Hard competitive learning, also known as learning vector quantization, is an algorithm for market segmentation that is different from the standard k-means algorithm in how segments are extracted. It randomly selects one consumer and moves their closest segment representative a small step in the direction of the randomly chosen consumer, rather than using all consumers in the data set to determine new segment representatives at each iteration. This procedural difference can lead to different segmentation solutions and hard competitive learning may find the globally optimal market segmentation solution while k-means gets stuck in a local optimum. An application of hard competitive learning in market segmentation analysis can be found in Boztug and Reutterer (2008), where it is used for segment-specific market basket analysis. Hard competitive learning can be computed in R using the function cclust(x, k, method = "hardcl") from package flexclust.

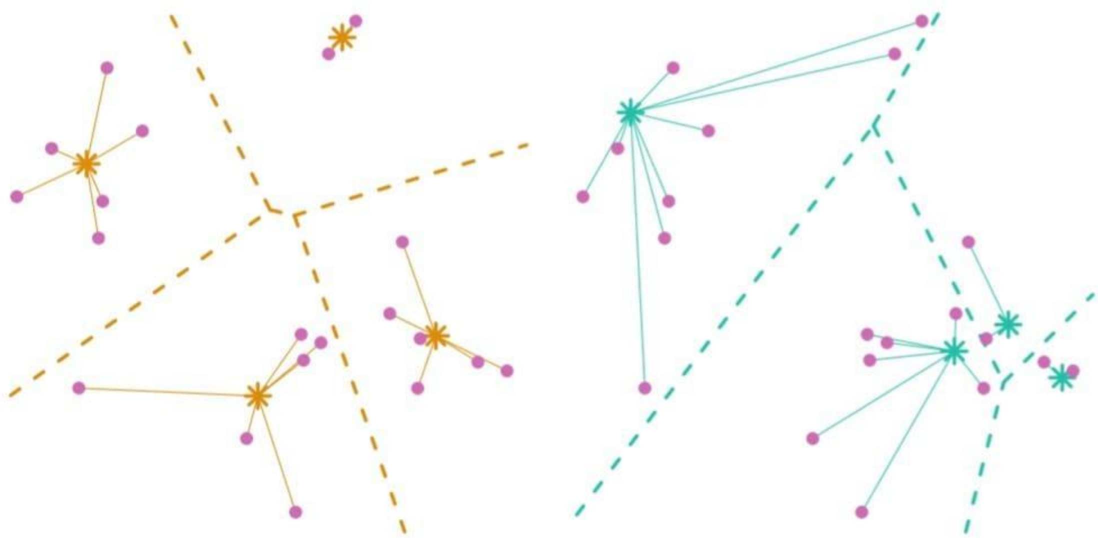


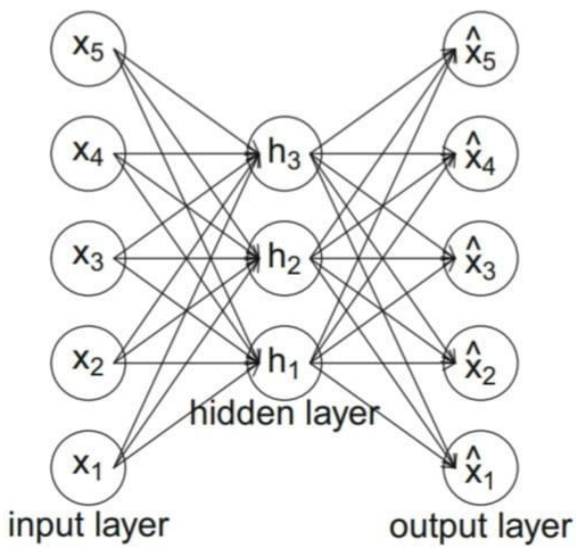
Fig. 7.15 Examples of good (*left*) and bad (*right*) starting points for *k*-means clustering

Neural Networks

Auto-encoding neural networks are a family of algorithms for cluster analysis that use a single hidden layer perceptron as the most popular method. The network has three layers: input, hidden, and output. The input layer takes the data as input, the output layer gives the response of the network, and the hidden layer has no connections to the outside of the network. The values of the nodes in the hidden layer are weighted linear combinations of the inputs for a non-linear function, and the outputs are weighted combinations of the hidden nodes. The method is mathematically different from all cluster methods presented so far, and it has been used in a marketing context by Natter (1999). Hruschka and Natter (1999) compare neural networks and k-means.

$$h_j = f_j\,(\textstyle\sum i=1 \text{ to } 5\ \alpha_{ij}x_i)$$

Fig. 7.18 Schematic representation of an auto-encoding neural network with one hidden layer



Hybrid Approaches

Hybrid segmentation approaches combine hierarchical and partitioning algorithms to compensate for their weaknesses. Hierarchical clustering algorithms allow for an unspecified number of market segments and visualisation, but require substantial memory capacity. Partitioning clustering algorithms have minimal memory requirements but require a specified number of segments and do not allow for tracking changes in segment membership. Hybrid approaches use partitioning algorithms to extract a larger number of segments initially, then retain only the segment centers and sizes for input into hierarchical clustering. This reduces the data set size and enables decision-making on the number of segments to extract.

Two-step Clustering

The two-step clustering procedure, implemented in IBM SPSS, involves a partitioning step followed by a hierarchical step. In the first step, k-means is used to extract a large number of clusters, from which representative members are retained to reduce the data set size. In the second step, hierarchical clustering is applied to the representatives to extract the desired number of segments. The original data is then linked to the segmentation solution using function twoStep() from package MSA. This procedure has been used in various application areas, and can be demonstrated using R commands on an artificial mobile phone data set. The resulting plot shows the correct segmentation solution extracted from the data.

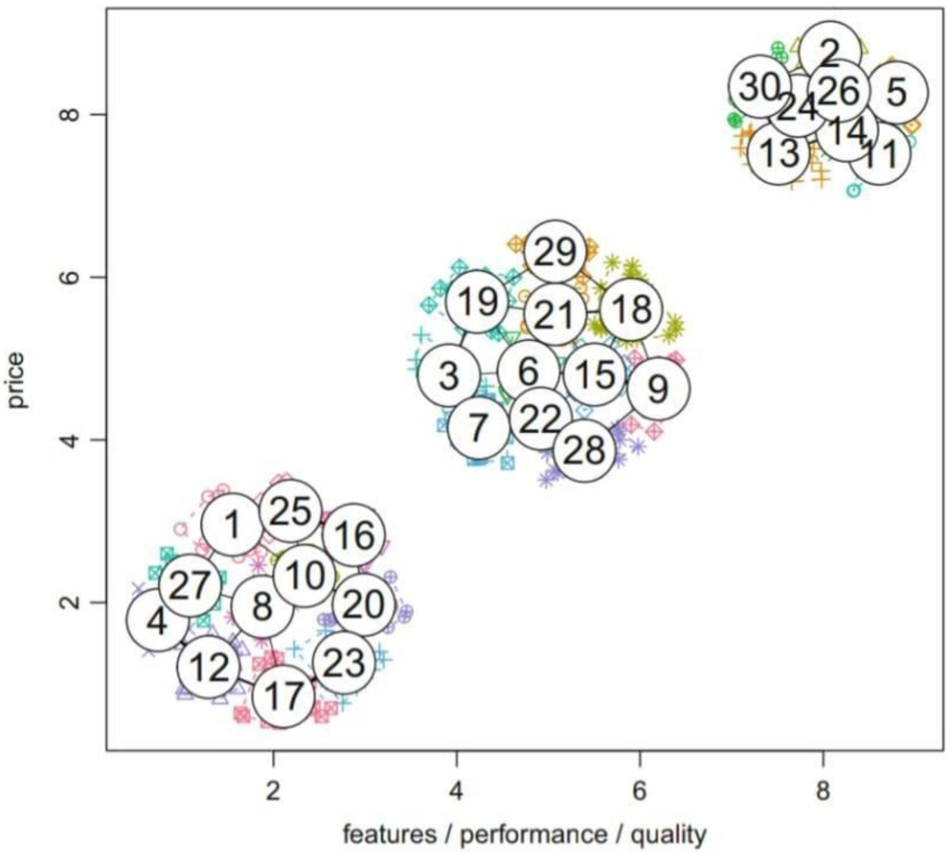


Fig. 7.19 *k*-means clustering of the artificial mobile phone data set into 30 clusters

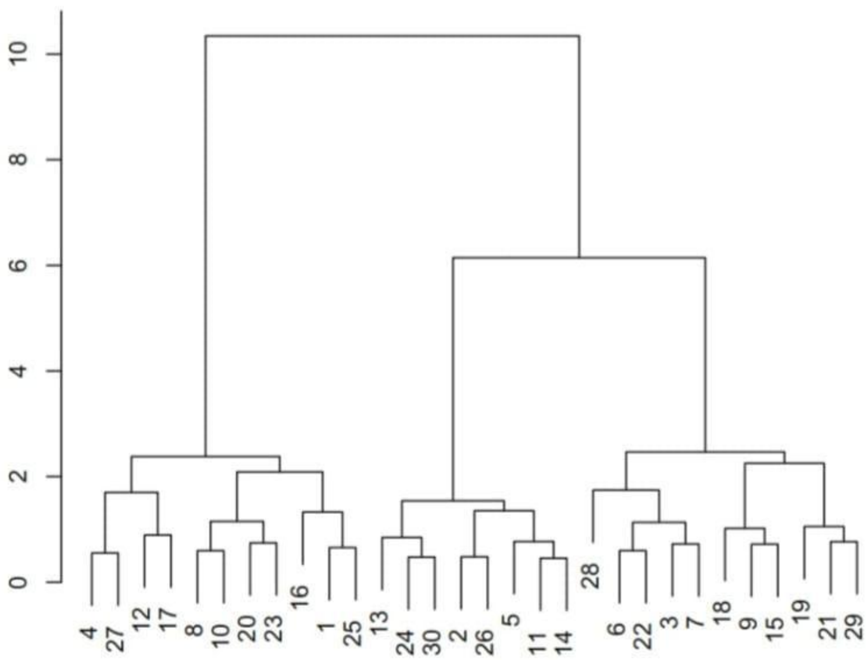


Fig. 7.20 Hierarchical clustering of the 30 *k*-means cluster centres of the artificial mobile phone data set

Bagged Clustering

Bagged clustering is a segmentation method that combines partitioning and hierarchical clustering algorithms with bootstrapping to identify market segments. The method involves creating multiple bootstrap samples from the original data set and then performing repeated partitioning clustering analyses to generate cluster centroids. These centroids are used to create a derived data set that is then subjected to hierarchical clustering analysis, resulting in a dendrogram that can provide clues about the number of market segments to extract. The final segmentation solution is determined by selecting a cut point for the dendogram and assigning each original observation to the closest market segment. The method has been successfully applied to tourism data, including a data set containing responses from 2961 tourists surveyed as part of the Austrian National Guest Survey in winter 1997/1998. The data set contains 27 binary segmentation variables related to winter vacation activities, and the marketing challenge is to identify tourist market segments based on these activities.

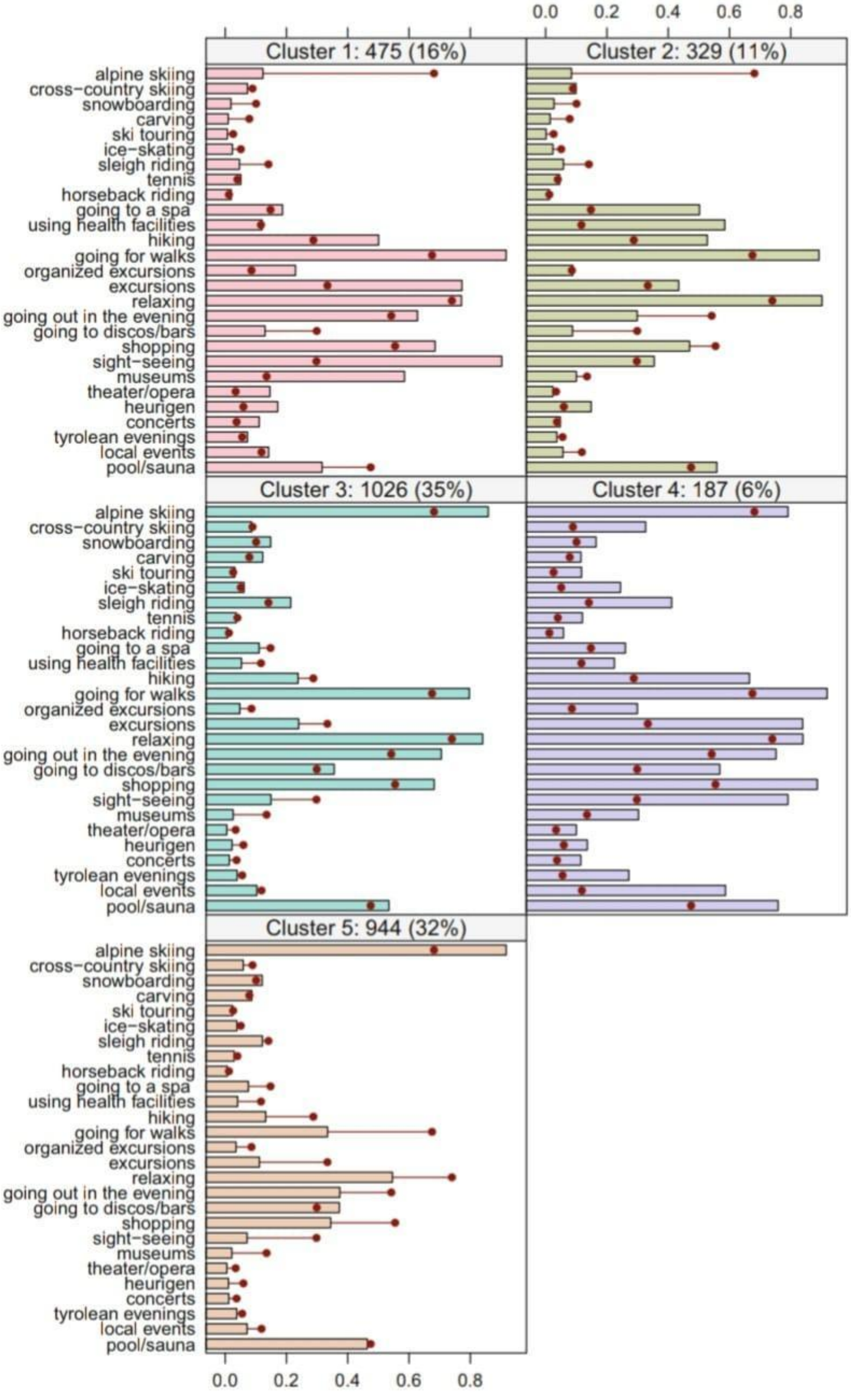


Fig. 7.22 Bar chart of cluster means from bagged cluster analysis of the winter vacation activities data set

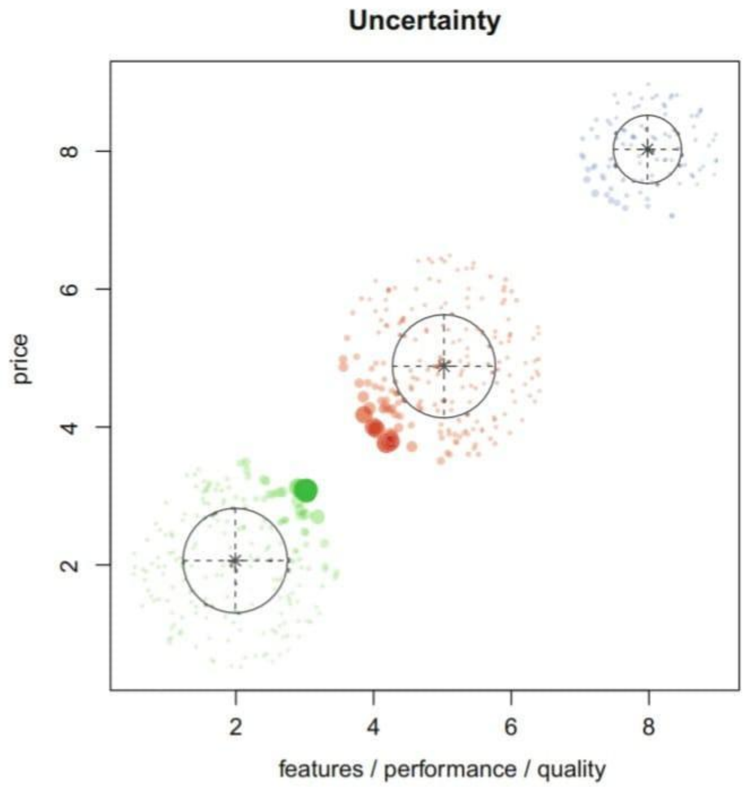
Model Based Methods

The article discusses the use of model-based methods as an alternative to distance-based methods in market segmentation analysis. The authors argue that using a range of extraction methods is helpful in determining the most suitable approach for the data at hand. Model-based methods do not rely on similarities or distances but instead assume that each market segment has a certain size and specific characteristics. These properties are not known in advance but are determined empirically using the data. The authors suggest that model-based methods offer a genuinely alternative extraction technique and may prove to be influential in marketing research.

Normal Distribution

The article discusses the use of a mixture of multivariate normal distributions as a popular finite mixture model for market segmentation analysis when dealing with metric data. This model is useful in cases where variables are correlated, and the multivariate normal distribution can model covariance between variables. The model has two sets of parameters: mean and variance. For each segment, there is a segment-specific mean vector and covariance matrix that contains the variances and covariances between the segmentation variables. The segment-specific parameters are the combination of the mean vector and covariance matrix, resulting in a total number of parameters to estimate of $p + p(p + 1)/2$.

Fig. 7.24 Uncertainty plot of the mixture of normal distributions for the artificial mobile phone data set



Binary Distribution

Finite mixtures of binary distributions, also known as latent class models or latent class analysis, are used for binary data where the segmentation variables are binary (0 or 1). This model assumes that different groups of respondents have different probabilities of undertaking certain activities, leading to negative correlation between certain variables in the overall data set. For example, some respondents may be interested in alpine skiing but not in sight-seeing, while others may be interested in sight-seeing but not in skiing. The mixture model allows us to identify these groups of respondents and their preferences.

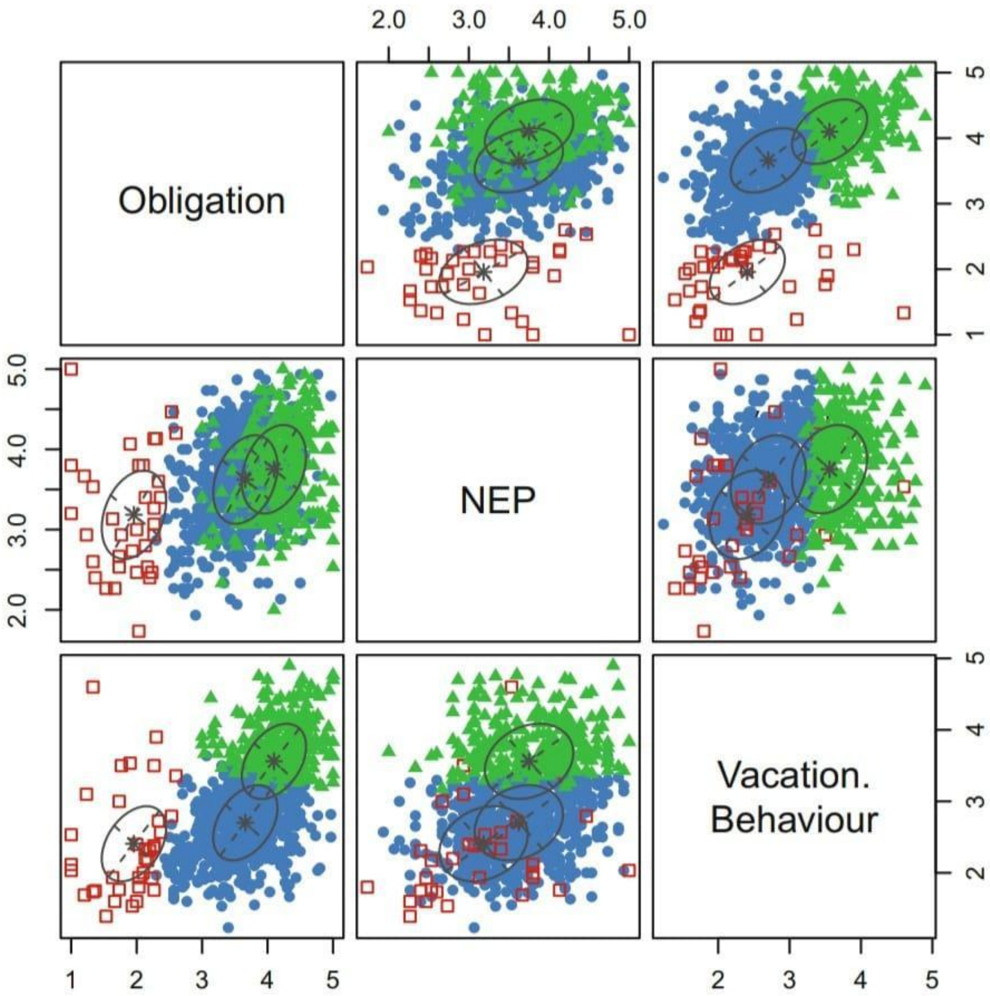


Fig. 7.29 Classification plot of the mixture of normal distributions for the Australian travel motives data set selected using the BIC among the models with identical covariance matrices across segments

Algorithms with Integrated Variable Selection

Biclustering Algorithms

Biclustering is a technique that simultaneously clusters both consumers and variables to identify groups of consumers who share common characteristics for a subset of variables. This approach is particularly useful for binary data such as genetic and proteomic data where traditional clustering algorithms are not effective due to the large number of variables and noisy data. Biclustering has experienced a big revival in recent years, and several popular biclustering algorithms exist with different ways of defining a bicluster. The goal of biclustering is to identify large groups of consumers who have as many variables in common as possible.

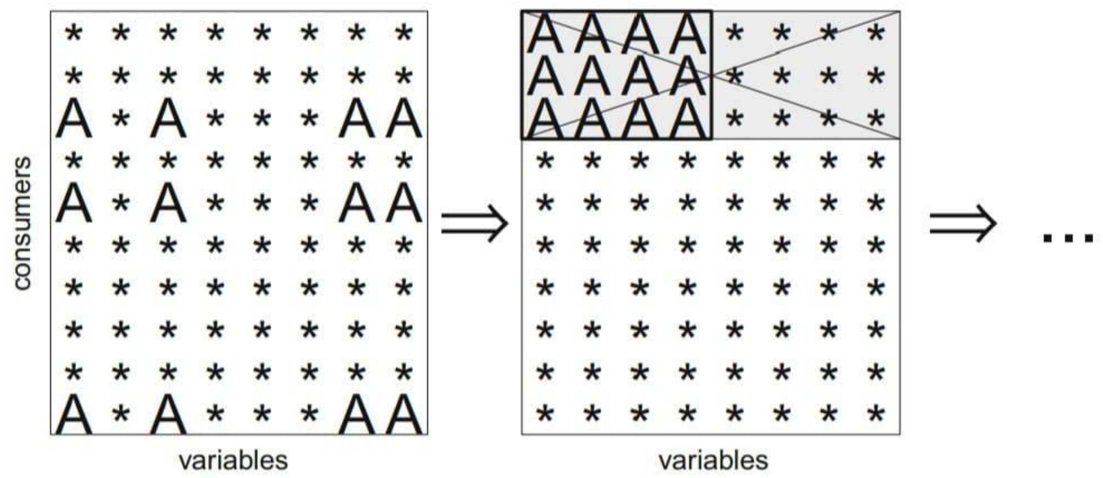


Fig. 7.35 Biclustering with constant pattern

Variable Selection Procedure for Clustering Binary Data (VSBD)

Brusco (2004) proposed the VSBD method for clustering binary datasets, based on the k-means algorithm. The method identifies the best small subset of variables for clustering, using the within-cluster sum-of-squares criterion. It then adds additional variables one by one, selecting the one leading to the smallest increase in the within-cluster sum-of-squares criterion, until a threshold is reached. The number of segments k has to be specified in advance, and can be selected using the Ratkowsky and Lance index. Brusco recommends a large number of random initialisations for the k-means algorithm, but using the more efficient Hartigan-Wong algorithm allows for fewer initialisations.

The algorithm works as follows:

Step 1: Select only a subset of observations with size $\varphi \in (0, 1]$ times the size of the original data set. Brusco (2004) suggests to use $\varphi = 1$ if the original data set contains less than 500 observations, $0.2 \leq \varphi \leq$

0.3 if the number of observations is between 500 and 2000 and φ = 0.1 if the number of observations is at least 2000.

Step 2: For a given number of variables V, perform an exhaustive search for the set of V variables that leads to the smallest within-cluster sum-of-squares criterion. The value for V needs to be selected small for the exhaustive search to be computationally feasible. Brusco (2004) suggests using V = 4, but smaller or larger values may be required depending on the number of clusters k, and the number of variables p. The higher the number of clusters, the larger V should be to capture the more complex clustering structure. The higher p, the smaller V needs to be to make the exhaustive search computationally feasible.

Step 3: Among the remaining variables, determine the variable leading to the smallest increase in the within-cluster sum-of-squares value if added to the set of segmentation variables.

Step 4: Add this variable if the increase in within-cluster sum-of-squares is smaller than the threshold. The threshold is δ times the number of observations in the subset divided by 4. δ needs to be in [0, 1]. Brusco (2004) suggests a default δ value of 0.5.

Variable Reduction: Factor-Cluster Analysis

The article discusses the use of factor-cluster analysis in market segmentation, which involves factor analyzing segmentation variables and using the resulting factor scores to extract market segments. The approach may be conceptually legitimate in cases where the data comes from validated psychological test batteries, but it is more commonly used when the number of segmentation variables is too high relative to the sample size. The article suggests that a sample size of at least 100 times the number of segmentation variables is needed for reliable results, but many studies use fewer variables and smaller sample sizes.

Data Structure Analysis

Cluster Indices

o make critical decisions in market segmentation analysis, data analysts need guidance, and cluster indices provide such guidance. Two types of cluster indices exist, namely internal and external cluster indices. Internal cluster indices use information contained in a single segmentation solution, while external cluster indices require another segmentation as additional input. The most commonly used measures of similarity of two market segmentation solutions are the Jaccard index, the Rand index and the adjusted Rand index. The former measures the similarity between two sets of segment memberships, while the latter two measure the agreement between two sets of segmentations.

Internal Cluster Indices

Internal cluster indices are used to evaluate the compactness and separability of market segments in a single segmentation solution. They require a distance measure between observations, a segment representative or centroid, and a representative for the complete data set. One common internal cluster index is the sum of within-cluster distances, which measures the compactness of clusters. The scree plot is a commonly used graph to select the number of market segments based on this index, and an elbow in the plot indicates a good segmentation solution. However, consumer data may not always have a distinct elbow in the scree plot, making it challenging to determine the optimal number of segments. Another variation of this index is the Ball-Hall index, which divides the sum of within-cluster distances by the number of segments to correct for the monotonous decrease of the internal cluster index with increasing numbers of segments.

$$W_k = \sum h=1 \text{ to } k \sum x \in S_h \, d(x, c_h)$$

External Cluster Indices

The passage discusses the evaluation of market segmentation solutions using external cluster indices, which require additional information beyond the original data. While the true segment structure is rarely known for consumer data, a repeated calculation using different clustering algorithms or variations of the original data can serve as external information. However, a problem with comparing segmentation solutions is label switching, where the labels of the segments are arbitrary and can vary between solutions. To address this, the focus should be on whether pairs of consumers are assigned to the same segments repeatedly, rather than on the individual segment labels. The passage outlines the four possible outcomes when comparing two market segmentation solutions for any two consumers.

• a: Both consumers are assigned to the same segment twice.

• b: The two consumers are in the same segment in P1, but not in P2.

• c: The two consumers are in the same segment in P2, but not in P1.

• d: The two consumers are assigned to different market segments twice.

Global Stability Analysis

Resampling methods can provide insight into the stability of market segmentation solutions across repeated calculations. To assess the global stability of a segmentation solution, several new data sets are generated using resampling methods, and a number of segmentation solutions are extracted. The stability of the segmentation solutions across repeated calculations is compared, and the solution that can best be replicated is chosen. Resampling methods can be valuable in market segmentation analysis because consumer data rarely contain distinct, well-separated market segments. There are three possible scenarios for consumer data: natural segments exist, data is entirely unstructured, or data lacks distinct, well-separated natural clusters. Global stability analysis helps determine which scenario applies to any given data set. The problem of sample randomness can be addressed by dividing the sample of respondents into subsamples and extracting market segments independently for each of the subsamples.

Segment Level Stability Analysis

Segment Level Stability Within Solutions(SLS_W)

The concept of segment level stability within solutions (SLSW) proposed by Dolnicar and Leisch (2017) assesses the stability of individual segments in a market segmentation solution, rather than the entire solution as a whole. This approach allows for the detection of highly stable segments, such as potential niche markets, even in solutions where other segments may be unstable. SLSW is calculated by repeatedly drawing bootstrap samples and calculating segmentation solutions independently for each sample, then determining the maximum agreement across all calculations using a method proposed by Hennig (2007). The authors demonstrate the procedure using an artificial mobile phone dataset and find that if the correct number of segments (three) are extracted, SLSW is high, but if more than three segments are extracted, some segments may have low SLSW due to splitting up of a larger natural segment. The SLSW approach is useful for organizations that only need to target one suitable market segment for survival and competitive advantage.

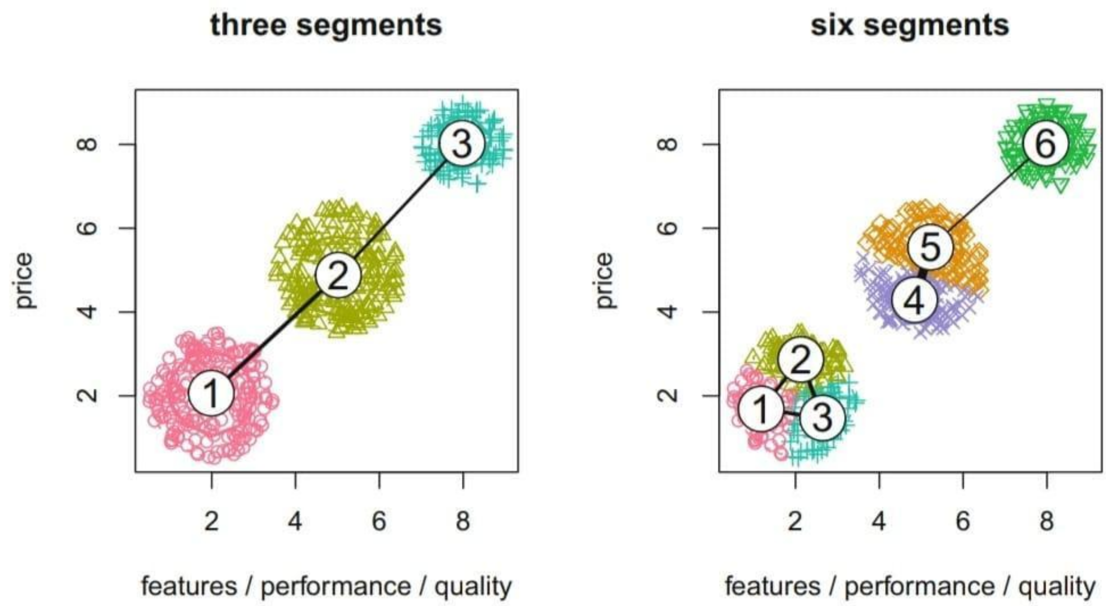


Fig. 7.41 Artificial mobile phone data set with three and six segments extracted

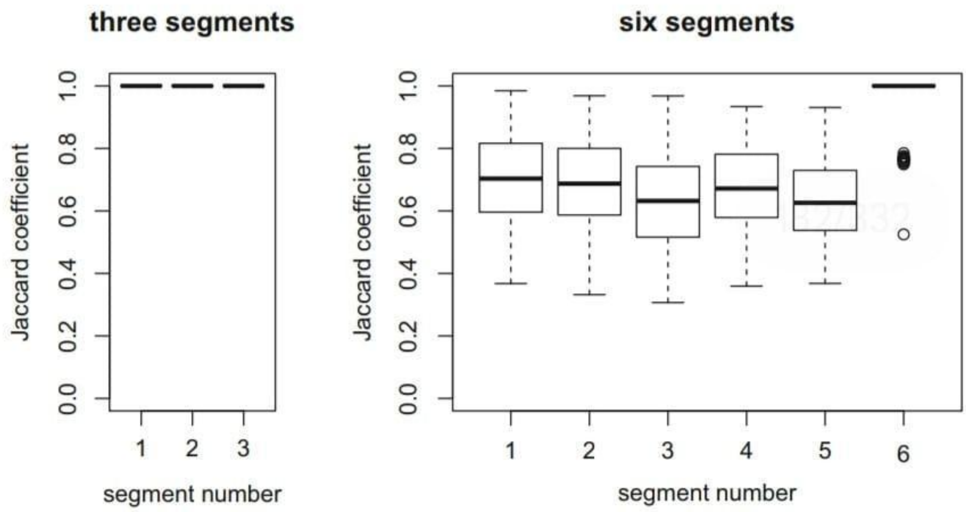


Fig. 7.42 Segment level stability within solutions (SLS_W) plot for the artificial mobile phone data set with three and six segments extracted

Segment Level Stability Across Solutions (SLS_4)

The segment level stability across solutions (SLSA) criterion is used to evaluate the stability of market segments across different segmentation solutions. High values of SLSA indicate natural segments that exist in the data, while low values suggest artificial segments created during the segmentation process. To calculate SLSA, a series of partitions with different numbers of segments are analyzed, and the similarity between segments in neighboring partitions is identified by relabeling the segments. The SLSA plot shows the development of each segment across different segmentation solutions, and thick lines indicate stubborn market segments that re-occur across solutions, while branching lines suggest changing segment membership and the possibility of artificially created segments. The SLSA criterion helps data analysts identify natural segments in the data, which are more attractive to organizations as they do not require managerial judgement for their creation.

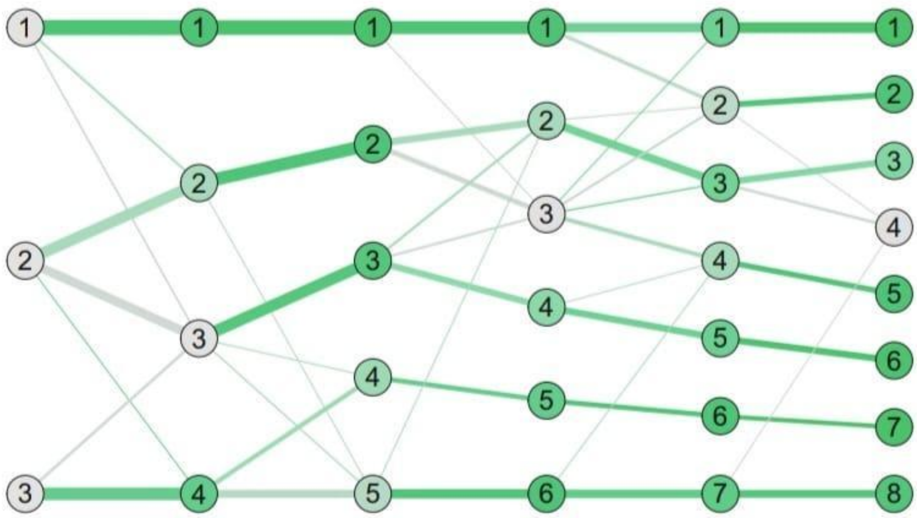


Fig. 7.45 Segment level stability across solutions (SLS_A) plot for the Australian travel motives data set for three to eight segments

Market Segment Analysis

Krishnendu Barman

1. Market Segments

1.1 Stratagic and tactical marketing:

The purpose of marketing is to match the genuine needs and desires of consumers with the offers of suppliers particularly suited to satisfy those needs and desires. This matching process benefits consumers and suppliers, and drives an organisation's marketing planning process. Marketing planning is a logical sequence and a series of activities leading to the setting of marketing objectives and the formulation of plans to achieving them (McDonald and Wilson 2011, p. 24). A marketing plan consists of two components: a strategic and a tactical marketing plan. The strategic plan outlines the long-term direction of an organisation, but does not provide much detail on shortterm marketing action required to move in this
long-term direction. The tactical marketing plan does the opposite. It translates the long-term strategic plan into detailed instructions for short-term marketing action. The strategic marketing plan states where the organisation wants to go and why. The tactical marketing plan contains instructions on what needs to be done to get there.

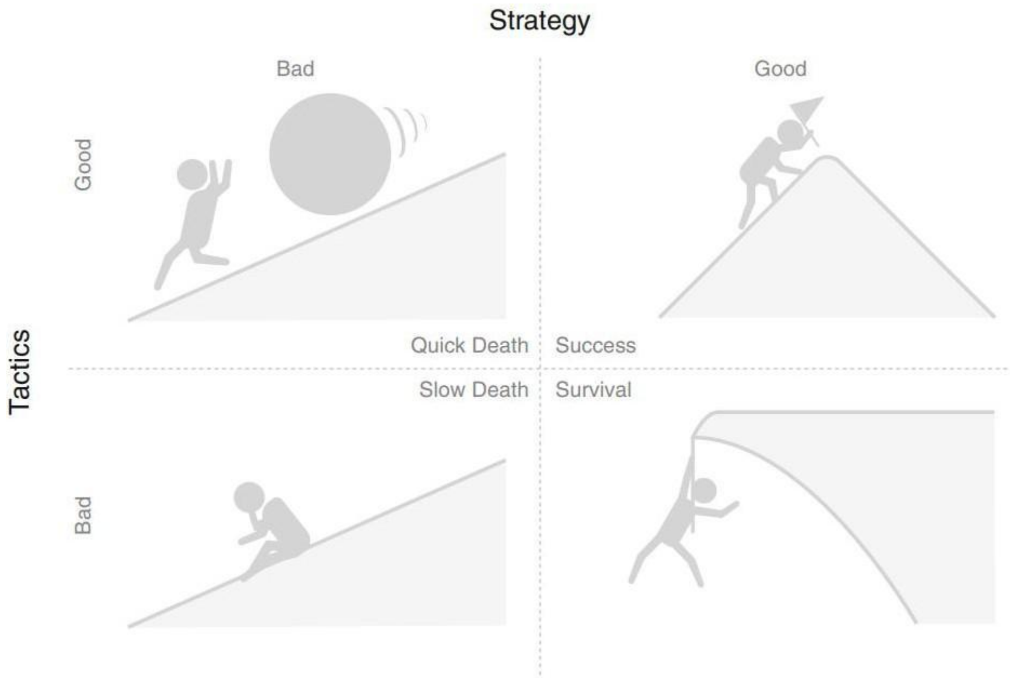


Fig 1 : The asymmetry of strategic and tactical marketing. (Modified from McDonald and Morris 1987)

The preparation for a mountain climbing expedition is similar to the development of an organisational marketing plan. The strategic marketing plan typically identifies consumer needs and desires, strengths and weaknesses internal to the organisation, and external opportunities and threats the organisation may face. A SWOT analysis outlines one side of the matching process: what the supplier is particularly suitable to offer consumers. Market research is used to explore what consumers need or desire, and two key decisions have to be made as part of the strategic marketing planning process: which consumers to focus on (segmentation and targeting) and which image of the organisation to create in the market (positioning). These decisions are critical because they determine the long-term direction of the organisation and cannot easily be reversed.

The importance of strategic and tactical marketing for organisational success is asymmetric, with strategic marketing being responsible for identifying the most suitable mountain to climb, while tactical marketing is responsible for the equipment. Good strategic marketing combined with good tactical marketing leads to the best possible outcome, while bad tactical marketing combined with good strategic marketing leads to failure. Good tactical marketing can never compensate for bad strategic marketing, and strategic marketing is the foundation of organisational success.

1.2Definitions of Market Segmentation:

Market segmentation is a decision-making tool for the marketing manager in selecting a target market for a given product and designing an appropriate marketing mix. It is essential for marketing success, as the most successful firms drive their businesses based on segmentation. Smith (1956) was the first to propose the use of segmentation as a marketing strategy, and it sits between the two extreme views that all objects are unique and inviolable and the population is homogeneous. Grey Advertising Inc. defines market segmentation as cutting markets into slices, with consumers belonging to the same market segments being very similar to one another with respect to the consumer characteristics deemed critical by management. The ideal market segmentation situation is that where three market segments exist: a small segment characterised by wanting many mobile telephone features, a large segment containing consumers who desire the exact opposite, and a large segment in the middle containing members who want a mid-range phone at a midrange price. This example illustrates Smith's definition of market segmentation, with each segment representing one homogeneous market within a larger heterogeneous market. Market segmentation is essential for organisational success.

For example : - Market segmentation is essential for organisational success, as a mobile phone company attempting to offer one mobile phone to the entire market is unlikely to satisfy the needs of each segment. Instead, tactical marketing efforts may be wasted if the company fails to cater for any of the homogeneous market segments. Selecting one market segment, such as the high-end, high-price segment, is more likely to lead to both short-term sales and long-term positioning.

A concentrated market strategy is an approach that focuses on satisfying the needs of one market segment, but comes with higher risk. An alternative approach is to pursue a differentiated market strategy, where all aspects of the marketing mix are customised for each of the three target segments. This approach is suitable in mature markets where consumers are capable of differentiating between alternative products. When an organisation decides not to use market segmentation, it is effectively choosing to pursue an undifferentiated market strategy. Examples of undifferentiated marketing include petrol and white bread. This approach may be viable for resource-rich organisations, or in cases where a new product is introduced.

1.3 The Benefits of market segments :

Market segmentation has a number of advantages, including the ability for businesses to reflect on what they are particularly excellent at and acquire insights into what customers desire. It also results in real benefits such as a greater understanding of consumer differences, which enhances the match between organisational strengths and consumer demands. Being able to cater to a specific specialised sector, such as micromarketing or hyper-segmentation, might provide a long-term competitive advantage. With the advent of eCommerce and the usage of sophisticated consumer database, finer segmentation tactics are becoming more realistic.

Market segmentation is important for small organisations to focus on satisfying the needs of a small group of consumers, as it allows direct sales efforts to be targeted at groups of consumers rather than each consumer individually. It can also contribute to team building and improve communication and information sharing across organisational units.

1.4 Cost of market segmentation :

Market segmentation requires a substantial investment by the organisation, requiring a large amount of time and resources to conduct a thorough analysis and develop and implement a customised marketing mix. It also requires an ongoing commitment of resources to evaluate the success of the strategy and monitor market dynamics. The upfront investment is substantial, but if not implemented well, it can lead to a waste of resources and a failed strategy can lead to substantial expenses generating no additional return, instead disenfranchising staff involved in the segmentation exercise.

It is for this very reason, that an organisation must make an informed decision about whether or not to embark on the long journey of market segmentation analysis, and the even longer journey of pursuing a market segmentation strategy.

2 Market Segmentation Analysis

2.1 The Layers of Market Segmentation Analysis :

This is the process of grouping consumers into naturally existing or artificially created segments of consumers who share similar product preferences or characteristics.

The first two layer of Market segmentation analysis is an exploratory process that requires both a data analyst and a user to be involved. The user is the person or department in the organisation that will use the results to develop a marketing plan. The most important details in this context are that the grouping of consumers is of the highest quality, and that the statistical segment extraction process at the core of market segmentation analysis cannot compensate for bad data. Additionally, the data needs to be explored to gain preliminary insight into the nature of the market segmentation study, and each segment needs to be profiled and described in detail. Profiling and describing segments help users understand each of the segments, and select which one(s) to target. Finally, profiling and describing segments inform the development of the customised marketing mix.

The third layer of market segmentation analysis includes nontechnical tasks that represent organisational implementation issues and why should we follow or not follow the first and second layers. These tasks must be converted into strategic marketing decisions and tactical marketing action to complete the analysis.

An organisation must assess whether implementing a market segmentation strategy will lead to market opportunities and be willing to commit to it. User input is essential at the data collection stage to ensure relevant information is captured. After segment extraction, users need to assess resulting market segments or market segmentation solutions, select one or more target segments, and develop a marketing plan for those market segments. This selection is driven by the strengths and opportunities of the organisation and their alignment with the key needs of the market segments.

2.2 Approaches to Market Segmentation Analysis:

There is no one single approaches to Market Segmentation Analysis. We must be diversified. In this study Market segmentation analysis is systematised in two ways: one based on the extent to which the organisation is willing or able to make changes to their current approach, and the other based on the nature of the segmentation variable or variables used in the analysis.

2.3 Based on Organisational Constraints:

Dibb and Simkin (2008) distinguish three approaches to market segmentation: quantitative survey-based, creation of segments from existing consumer classifications, and emergence of segments from qualitative research. The most radical approach is segment revolution, while segment evolution is like refining an existing sandcastle. The least radical approach is a random discovery, like a mutation, which has the potential to harvest the benefits of segmentation.

The segment revolution or quantitative survey-based segmentation approach is the most common approach for market segmentation analysis. It assumes that the organisation is willing to start from scratch and start with an open mind. If the segmentation analysis reveals a promising niche segment or a set of market segments to target, the organisation must develop an entirely new marketing plan in view of those findings.

Market segmentation analysis is often not viable due to unwillingness or inability of an organisation to change, but it does not have to be abandoned altogether. Other, less radical approaches are available, such as creating segments from currently targeted sectors and segments. Dibb and Simkin (2008) offer a proforma to guide organisations through this process.

The third approach is that of exploratory research pointing to segments, where market segments are found as part of an exploratory research process. In times of big data, segment mutation may result from data mining of streams of data, rather than qualitative research. Continuous tracking of the nature of market segments in large streams of data can be used to check if market structure has changed in ways that make it necessary to adapt segmentation strategy to ensure organisational survival and prosperity.

2.4 Based on the Choice of (the) Segmentation Variable(s):

Segmentation approaches use the nature of consumer characteristics to extract market segments. One single piece of information about consumers (one segmentation variable) is used, such as age. In other cases, multiple pieces of information (multiple segmentation variables) are important, such as consumers' expenditure patterns. For example, the information collection for a tourism company the commonly used segmentation variables are provided in Table 2.1.

When one single segmentation variable is used, the segmentation approach is referred to as a priori (Mazanec 2000), convenience-group (Lilien and Rangaswamy 2003) or commonsense market segmentation (Dolnicar 2004). Morritt (2007) describes this approach as one that is created without the benefit of primary market research. Common sense segmentation implies that users apply their common sense to choose their target segment, while convenience-group segmentation is chosen for the convenience of serving them. The aim of the segmentation analysis is not to identify the key defining characteristic of the segment, but to gain deeper insight into the nature of the segments.

Commonsense segmentation is the use of powerful segmentation variables to identify consumers who have distinct profiles with respect to descriptor variables. It is an efficient approach because it is simpler and fewer mistakes can occur. Lilien and Rangaswamy (2003) view this approach as reactive, whereas in data-driven segmentation, the nature of the resulting market segments is not known until after the data analysis has been conducted. The goal of data-driven segmentation is to explore different market segments that can be extracted using the segmentation variables chosen and to develop a detailed profile and description of the segment(s) selected for targeting. The excerpt further provides examples of commonly used segmentation variables and approaches to market segmentation studies. In reality, market segmentation studies usually involve various combinations of these approaches.

Steps of market segment analysis:

3 Step 1: Deciding (not) to Segment

3.1 Implications of Committing of Market Segment Analysis:

The passage discusses the potential advantages and drawbacks of pursuing a market segmentation strategy in organizations. The key implication of market segmentation is the need for a long-term commitment to the strategy. The organization must be willing and able to make substantial changes and investments, including the development of new products, modification of existing products, changes in pricing and distribution channels, and communication with the market. These changes can influence the internal structure of the organization, requiring adjustments to align with the needs of different market segments. To maximize the benefits of market segmentation, organizations should organize around market segments, rather than products, and form strategic business units in charge of segments to ensure ongoing focus on their changing needs.

The decision to pursue market segmentation must be made at the highest executive level and systematically communicated and reinforced across all organizational levels and units due to the significant long-term commitment required. The cost of implementing a segmentation strategy must be justified by the expected increase in sales.

3.2 Implementation Barrier :

The text discusses the implementation barriers that can prevent the successful roll-out of a market segmentation strategy. These barriers are categorized into several groups. The first group is related to senior management and includes the lack of leadership, pro-active championing, commitment, and involvement in the market segmentation process. The second group is related to organizational culture and includes lack of market or consumer orientation, resistance to change and new ideas, lack of creative thinking, bad communication, and lack of sharing of information and insights across organizational units, short-term thinking, unwillingness to make changes, and office politics. Another potential problem is the lack of training.

The lack of a formal marketing function or at least a qualified marketing expert in the organization is another obstacle that can prevent the successful implementation of market segmentation. Other obstacles may include objective restrictions faced by the organization, such as lack of financial resources or the inability to make the structural changes required. Process-related barriers include not having clarified the objectives of the market segmentation exercise, lack of planning or bad planning, a lack of structured processes to guide the team through all steps of the market segmentation process, a lack of allocation of responsibilities, and time pressure.

To counteract these challenges, it recommends making market segmentation analysis easy to understand and presenting results in a way that facilitates interpretation by managers. Most of these barriers can be identified from the outset of a market segmentation study and proactively removed. If barriers cannot be removed, the option of abandoning the attempt of exploring market segmentation as a potential future strategy should be seriously considered. Finally, the text recommends that a resolute sense of purpose and dedication is required, tempered by patience and a willingness to appreciate the inevitable problems encountered in implementing the conclusions.

Step 1 Checklist:

This first checklist includes not only tasks, but also a series of questions which, if not answered in the affirmative, serve as knock-out criteria. For example: if an organisation is not market-oriented, even the finest of market segmentation analyses cannot be successfully implemented.

4 Specifying the Ideal Target Segment:

4.1 Segment Evaluation Criteria

This section of market segmentation analysis, which is the process of dividing a larger market into smaller groups of consumers who share similar needs and characteristics. The third layer of market segmentation analysis is dependent on user input and involvement throughout the process, not just at the beginning or end.

To produce useful results, the organization must make a major contribution to the market segmentation analysis, starting with conceptualizing the segment evaluation criteria in Step 2. The organization must determine two sets of criteria: knock-out criteria, which are essential and non-negotiable features of segments that the organization would consider targeting, and attractiveness criteria, which are used to evaluate the relative attractiveness of the remaining market segments.

The knock-out criteria automatically eliminate some of the available market segments, while attractiveness criteria are negotiated by the segmentation team and applied to determine the overall relative attractiveness of each market segment in Step 8. Members of the segmentation team need to select which attractiveness criteria they want to use to determine how attractive potential target segments are and assess the relative importance of each criterion to the organization.

Overall, the text emphasizes the importance of involving the user in most stages of market segmentation analysis, including conceptualizing segment evaluation criteria, to produce results that are useful to the organization.

4.2 Knock-Out Criteria:

- The segment must be homogeneous; members of the segment must be similar to one another.
- The segment must be distinct; members of the segment must be distinctly different from members of other segments.
- The segment must be large enough; the segment must contain enough consumers to make it worthwhile to spend extra money on customising the marketing mix for them.
- The segment must be matching the strengths of the organisation; the organisation must have the capability to satisfy segment members' needs.
- Members of the segment must be identifiable; it must be possible to spot them in the marketplace.
- The segment must be reachable; there has to be a way to get in touch with members of the segment in order to make the customised marketing mix accessible to them.

4.3 Attractiveness Criteria:

There is a range of segment attractiveness criteria, which are not binary and can be more or less attractive. The attractiveness across all criteria determines whether a market segment is selected as a target segment in Step 8 of market segmentation analysis.

4.4 Implementing a structured process:

The passage discusses the importance of following a structured process when assessing market segments for selection as target markets. It highlights the use of a segment evaluation plot, which is a popular approach for evaluating market segments based on their attractiveness and organisational competitiveness. The values for these criteria are determined by the segmentation team, and it is important to negotiate and agree on the factors that constitute them. A team of people should be involved in this process, including representatives from different organisational units, as each unit has a different perspective on the business and will be affected by the segmentation strategy.

The segment evaluation plot cannot be completed at the early stages of market segmentation analysis as no segments are available yet. However, selecting the attractiveness criteria for market segments at this stage ensures that all relevant information is captured during data collection and makes the task of selecting a target segment easier later on.

The market segmentation team should have a list of approximately six segment attractiveness criteria at the end of this step, with a weight attached to each criterion to indicate its importance to the organisation. The typical approach to weighting is to ask all team members to distribute 100 points across the criteria, which are then negotiated until agreement is reached. It is important to seek approval from the advisory committee, which contains representatives from multiple organisational units, bringing different perspectives to the challenge of specifying segment attractiveness criteria.

5 Colleting Data :

5.1 Segmentation variable:

the role of empirical data in market segmentation. Empirical data refers to data that has been gathered through observation or experiment, and is used to identify or create market segments.

The excerpt introduces the concept of a "segmentation variable," which is the variable in the empirical data that is used to split the sample into market segments. In commonsense segmentation, the segmentation variable is typically a single characteristic of the consumers in the sample, such as gender. The excerpt provides an example in 1st table in this section, where each row represents one consumer and each variable represents one characteristic of that consumer. In this example, the segmentation variable is gender, and the sample is split into a segment of women and a segment of men.

In addition to the segmentation variable, there are other personal characteristics available in the data, such as age, the number of vacations taken, and information about benefits sought when going on vacation. These characteristics are called "descriptor variables" and are used to describe the segments in detail. The excerpt notes that data-driven market segmentation is based on multiple segmentation variables, rather than just one. These variables are used to identify naturally existing or artificially created market segments that are useful to the organization. The excerpt provides an example in table 2, using the same data as in 1st .

Sociodemographics		Travel behaviour	Benefits sought				
gender	age	N° of vacations	relaxation	action	culture	explore	meet people
Female	34	2	1	0	1	0	1
Female	55	3	1	0	1	0	1
Female	68	1	0	1	1	0	0
Female	34	1	0	0	1	0	0
Female	22	0	1	0	1	1	1
Female	31	3	1	0	1	1	1
Male	87	2	1	0	1	0	1
Male	55	4	0	1	0	1	1
Male	43	0	0	1	0	1	0
Male	23	0	0	1	1	0	1
Male	19	3	0	1	1	0	1
Male	64	4	0	0	0	0	0
segmentation variable	descriptor variables						

Table 1: Gender as a possible segmentation variable in commonsense market segmentation

Sociodemographics		Travel behaviour	Benefits sought				
gender	age	N° of vacations	relaxation	action	culture	explore	meet people
Female	34	2	1	0	1	0	1
Female	55	3	1	0	1	0	1
Male	87	2	1	0	1	0	1
Female	68	1	0	1	1	0	0
Female	34	1	0	0	1	0	0
Female	22	0	1	0	1	1	1
Female	31	3	1	0	1	1	1
Male	55	4	0	1	0	1	1
Male	43	0	0	1	0	1	0
Male	23	0	0	1	1	0	1
Male	19	3	0	1	1	0	1
Male	64	4	0	0	0	0	0
descriptor variables			segmentation variables				

Table2:Segmentation variables in data-driven market segmentation

5.2 Segmentation Criterion:

Before segmenting a market, an organization must decide on a segmentation criterion, which refers to the nature of the information used for market segmentation. The most common criteria are geographic, socio-demographic, psychographic, and behavioral. However, choosing the best criterion to use requires prior knowledge about the market, and there are few guidelines to help make this decision. Bock and Uncles (2002) suggest that profitability, bargaining power, preferences for benefits or products, barriers to choice, and consumer interaction effects are the most relevant differences between consumers for market segmentation. Despite the different criteria available, the recommendation is to use the simplest possible approach that works for the product or service at the least possible cost.

Cahill (2006) emphasizes this point by saying that the organization should "Do the least you can." For example, if demographic segmentation will work, then use it; there is no need to choose a more sophisticated approach such as psychographic segmentation just because it is considered sexier.

5.2.1 Geographic Se1gmentation:

Geographic segmentation is the original criterion used for market segmentation, and it typically involves using a consumer's location of residence to form market segments. While it is a simple approach, it can be the most appropriate, as illustrated by examples such as national tourism organizations needing to use different languages to attract tourists from neighboring countries. Global companies like Amazon and IKEA also use geographic segmentation to offer country-specific information and pricing. The main advantage of geographic segmentation is that it is easy to assign each consumer to a geographic unit, making it easy to target communication messages and select communication channels to reach the selected geographic segments, such as local newspapers and TV stations.

Geographic segmentation has a key disadvantage because living in the same location does not necessarily mean that people share other relevant characteristics, such as their preferred product benefits. Even in cases where location is a factor, other criteria, such as socio-demographics, are usually the reason for similarities in product preferences. For example, people from the same country are likely to have different ideal holidays depending on various factors. Despite these shortcomings, there has been a recent revival in using geographic information as a segmentation variable in international market segmentation studies. However, this approach is challenging because the segmentation variable must be meaningful across all included regions, and surveys completed by respondents from different cultural backgrounds can introduce biases. An example of such a study is provided by Haverila (2013), who extracted market segments of mobile phone users among young customers across national borders.

5.2.2 Socio-Demographic Segmentation:

Socio-demographic segmentation is based on age, gender, income, and education, and can be useful in industries such as luxury goods, cosmetics, and retirement villages. However, socio-demographic criteria may not always be the cause for product preferences and may not provide sufficient market insight. Demographics explain about 5% of the variance in consumer behaviour, and values, tastes, and preferences are more useful for market segmentation.

5.2.3 Psychographic Segmentation:

Psychographic segmentation is when people are grouped according to their psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product. This approach is more complex than geographic or socio-demographic segmentation because it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest. Benefit segmentation and lifestyle segmentation are popular approaches to psychographic segmentation, which often use a number of different segmentation variables.

This approach is advantageous as it reflects the underlying reasons for differences in consumer behavior. However, determining segment memberships for consumers is complex, and the reliability and validity of empirical measures used to capture psychographic dimensions are crucial to the success of this approach. The use of travel motives as a basis for data-driven market segmentation in tourism is an example of the psychographic approach.

5.2.4 Behavioural Segmentation:

Behavioral criteria can include factors such as prior experience with a product, frequency of purchase, amount spent on purchasing the product, and information search behavior. One advantage of behavioral segmentation is that it allows for segmentation based on actual behavior, rather than self-reported or intended behavior, and can be more reflective of the underlying reasons for differences in consumer behavior. For example, tourists whose travel behavior is characterized by a preference for cultural experiences are more likely to undertake a cultural holiday at a destination that has ample cultural treasures for them to explore.Behavioral segmentation has been found to be superior to geographic variables in segmentation analyses, as reported by Moscardo et al. (2001). The disadvantage of behavioral approaches is that behavioral data may not always be readily available, especially for potential customers who have not previously purchased the product. However, when behavioral data is available, it can be a powerful tool for segmenting consumers based on the similarities that matter most.

5.3 Data from Survey Studies:

Most market segmentation analyses are based on survey data. Survey data is cheap and easy to collect, making it a feasible approach for any organisation. But survey data – as opposed to data obtained from observing actual behaviour – can be contaminated by a wide range of biases. Such biases can, in turn, negatively affect the quality of solutions derived from market segmentation analysis. A few key aspects that need to be considered when using survey data are discussed below.

5.3.1 Choice of Variables:

Careful selection of segmentation variables is crucial for high-quality market segmentation. In data-driven segmentation, all relevant variables should be included, while unnecessary ones must be avoided to prevent respondent fatigue and make the segmentation problem unnecessarily difficult. Noisy or masking variables, which divert the algorithm's attention away from critical information, can prevent algorithms from identifying the correct segmentation solution. They can arise from poorly developed survey questions or not carefully selecting segmentation variables. The issue of noisy variables can be avoided by asking all necessary questions and avoiding unnecessary or redundant questions. Conducting exploratory or qualitative research can provide insights about people's beliefs that survey research cannot offer, which can be included as answer options in a questionnaire. This two- stage process ensures that no critically important variables are left out. Redundant items in a survey are particularly problematic for market segmentation analysis because they interfere with most segment extraction algorithms' ability to identify correct market segmentation solutions.

5.3.2 Response Option:

The answer options provided to survey respondents determine the scale of data available for subsequent analyses. Binary or dichotomous data, generated from options with only two ways to answer, can be represented in a data set by 0s and 1s, which pose no difficulties for subsequent segmentation analysis. Nominal variables, which correspond to options allowing respondents to select an answer from a range of unordered categories, can be transformed into binary data by introducing a binary variable for each answer option. These options are important to consider as many data analytic techniques are based on distance measures, and not all survey response options are equally suitable for segmentation analysis.

When respondents are asked to indicate a number, such as age or nights stayed at a hotel, metric data is generated, which is well-suited for segmentation analysis as it allows any statistical procedure to be performed. However, the most commonly used response option in survey research is a limited number of ordered answer options larger than two, where respondents are asked to express their agreement with a series of statements using five or seven response options. This generates ordinal data where the options are ordered, but the distance between adjacent answer options is not clearly defined, making it difficult to apply standard distance measures to such data without making strong assumptions. Suitable distance measures for each scale level are discussed in detail in Step 5.

It suggests that using binary or metric response options is preferable to avoid complications with distance measures in segmentation analysis. Although ordinal scales are commonly used, using binary or metric response options instead is usually not a compromise. Visual analogue scales can also be used to capture fine nuances of responses. In many contexts, binary response options have been shown to outperform ordinal answer options, especially when formulated in a level-free way.

5.3.3 Response Styles:

Response biases in survey data and how they can impact segmentation results. Response biases are systematic tendencies of respondents to answer questions based on factors other than the content of the questions. These biases can lead to response styles, which affect segmentation results because they cannot be distinguished from genuine responses. The passage suggests that the risk of capturing response styles should be minimized when collecting data for segmentation purposes. If attractive market segments emerge with response patterns potentially caused by a response style, additional analyses are required to exclude this possibility or respondents affected by such a response style must be removed before targeting such a market segment.

5.3.4 Sample size:

Insufficient sample size can make it difficult to determine the correct number of market segments. Some studies recommend a minimum sample size of 2p or five times 2p, where p is the number of segmentation variables. Other studies suggest a sample size of at least 10 times the number of segmentation variables times the number of segments. Dolnicar et al. conducted simulation studies to test sample size requirements for algorithms to correctly identify segments. The adjusted Rand index is used to measure the correctness of segment recovery, with higher values indicating better alignment. In Fig. 2, the x-axis plots the sample size (ranging from 10 to 100 times the number of segmentation variables). The y-axis plots the effect of an increase in sample size on the adjusted Rand index. The higher the effect, the better the algorithm identified the correct market segmentation solution.

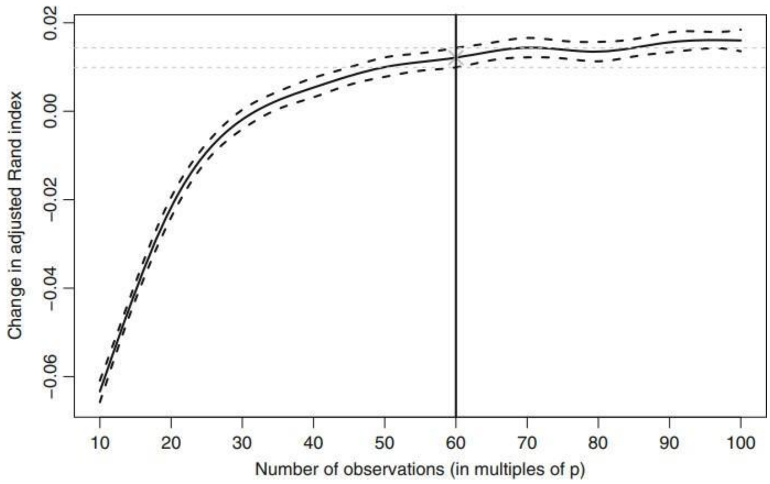


Fig 2 : Effect of sample size on the correctness of segment recovery in artificial data. (Modified from Dolnicar et al. 2014)

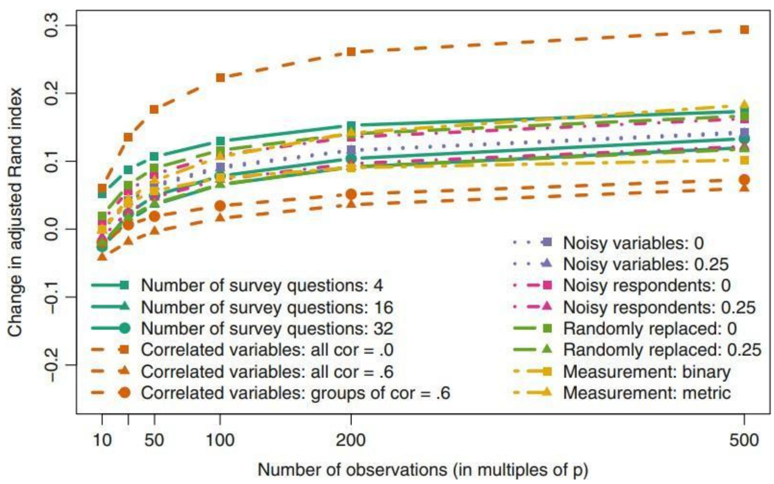


Fig 3 : Sample size requirements in dependence of market and data characteristics. (Modified from Dolnicar et al. 2016)

Attach fig

Increasing sample size improves the correctness of extracted segments, with the biggest improvement seen in very small samples. A sample size of at least 60 · p is recommended for typical survey data sets, while a sample size of at least 70 · p is recommended for more difficult scenarios. Market characteristics such as the number and size of segments, and the extent of overlap, can affect the difficulty of identifying correct segments. Survey data characteristics such as sampling error, response biases, low data quality, and item correlation can also affect segment recovery.

Fig 3 shows the results from this large-scale simulation study using artificial data. Again, the axes plot the sample size, and the effect of increasing sample size on the adjusted Rand index, respectively.

But how larger sample sizes improve an algorithm's ability to identify the correct market segmentation solution, but the extent of this improvement varies based on market and data characteristics. Using uncorrelated segmentation variables leads to good segment recovery, while correlation between variables cannot be well compensated for by increasing the sample size. The study recommends a sample size of at least 100 respondents for each segmentation variable and emphasizes the importance of collecting high-quality unbiased data for market segmentation analysis.

It can be concluded that data used in market segmentation analyses should –

- contain all necessary items;
- contain no unnecessary items;
- contain no correlated items;
- contain high-quality responses;
- be binary or metric;
- be free of response styles;
- include responses from a suitable sample given the aim of the segmentation study; and
- include a sufficient sample size given the number of segmentation variables (100 times the number of segmentation variables).

5.4 Data from Internal Sources:

Organizations have access to large amounts of internal data for market segmentation analysis, such as scanner data and online purchase data, which represent actual consumer behavior. This data is advantageous because it is automatically generated and requires no extra effort to collect. However, there is a danger of using internal data because it may be systematically biased by over-representing existing customers and not providing information about potential future customers with different consumption patterns. Also notes that consumer statements about their behavior or intentions can be affected by imperfect memory and response biases.

5.5 Data from Experimental Studies :

experimental data can be used as a source for market segmentation analysis, which can result from field or laboratory experiments. Examples include testing how people respond to certain advertisements or presenting consumers with carefully developed stimuli consisting of specific levels of specific product attributes to determine their preferences. The resulting information about consumer preferences and the impact of different attributes and attribute levels can be used as a segmentation criterion.

6 Step VI : Profiling Segments:

6.1 Identifying Key Characteristics of Market Segments:

The profiling step is necessary for data-driven market segmentation, but not for commonsense segmentation. Profiling involves identifying defining characteristics of market segments with respect to segmentation variables. It is important to inspect alternative segmentation solutions, especially if no natural segments exist in the data. Good profiling is critical for correct interpretation of the resulting segments, which is important for making good strategic marketing decisions.

Marketing managers often face in interpreting data-driven market segmentation solutions. Studies show that a significant percentage of marketing managers struggle to understand such solutions, with 65% reporting difficulties interpreting them correctly. Furthermore, 71% of the managers feel that segmentation analysis is like a black box. The article provides quotes from marketing managers illustrating how market segmentation results are often presented to them. Some of these are :

- . . . as a long report that usually contradicts the results
- . . . rarely with a clear Executive Summary
- . . . in a rushed slap hazard fashion with the attitude that 'leave the details to us' ...
- The result is usually arranged in numbers and percentages across a few (up to say 10) variables, but mostly insufficiently conclusive.
- ...report or spreadsheet...report with percentages
- . . . often meaningless information • In a PowerPoint presentation with a slick handout

1.1 Traditional Approaches to Profiling Market Segments:

Data-driven segmentation solutions are usually presented to users (clients, managers) in one of two ways: (1) as high level summaries simplifying segment characteristics to a point where they are misleadingly trivial, or (2) as large tables that provide, for each segment, exact percentages for each segmentation variable. Such tables are hard to interpret, and it is virtually impossible to get a quick overview of the key insights.

Sometimes – to deal with the size of this task – information is provided about the statistical significance of the difference between segments for each of the segmentation variables. This approach, however, is not statistically correct. Segment membership is directly derived from the segmentation variables, and segments are created in a way that makes them maximally different, thus not allowing to use standard statistical tests to assess the significance of differences.

6.2 Segment Profiling with Visualisations:

The importance of using graphics in data-driven market segmentation solutions is very handy. Currently, these solutions are presented using either highly simplified or complex tabular representations, which do not make optimal use of graphics. However, graphics play a crucial role in exploratory statistical analysis, such as cluster analysis, as they provide insights into the complex relationships between variables. Additionally, in times of big data, visualization offers a simple way of monitoring developments over time.

The article highlights the recommendations of McDonald and Dunbar and Lilien and Rangaswamy to use visualization techniques to make the results of a market segmentation analysis easier to interpret.

At last we can say that visualizations are useful in the data-driven market segmentation process to inspect one or more segments in detail and to assess the usefulness of a market segmentation solution. The process of segmenting data leads to a large number of alternative solutions, and selecting one of the possible solutions is a critical decision. Visualizations of solutions assist the data analyst and user with this task. Therefore, incorporating graphics in data-driven market segmentation solutions can lead to better insights and decision-making.

6.3 Identifying Defining Characteristics of Market Segments :

A good way to understand the defining characteristics of each segment is to produce a segment profile plot. The segment profile plot shows – for all segmentation variables – how each market segment differs from the overall sample.

Segmentation variables do not have to be displayed in the order they appear in the data set in figures and tables. If variables in the data set have a meaningful order, that order should be preserved. If, on the other hand, the order of variables is independent. It is beneficial to rearrange variables in order to improve visualisations. We can use colors and markers for our convenient explanation. As example:

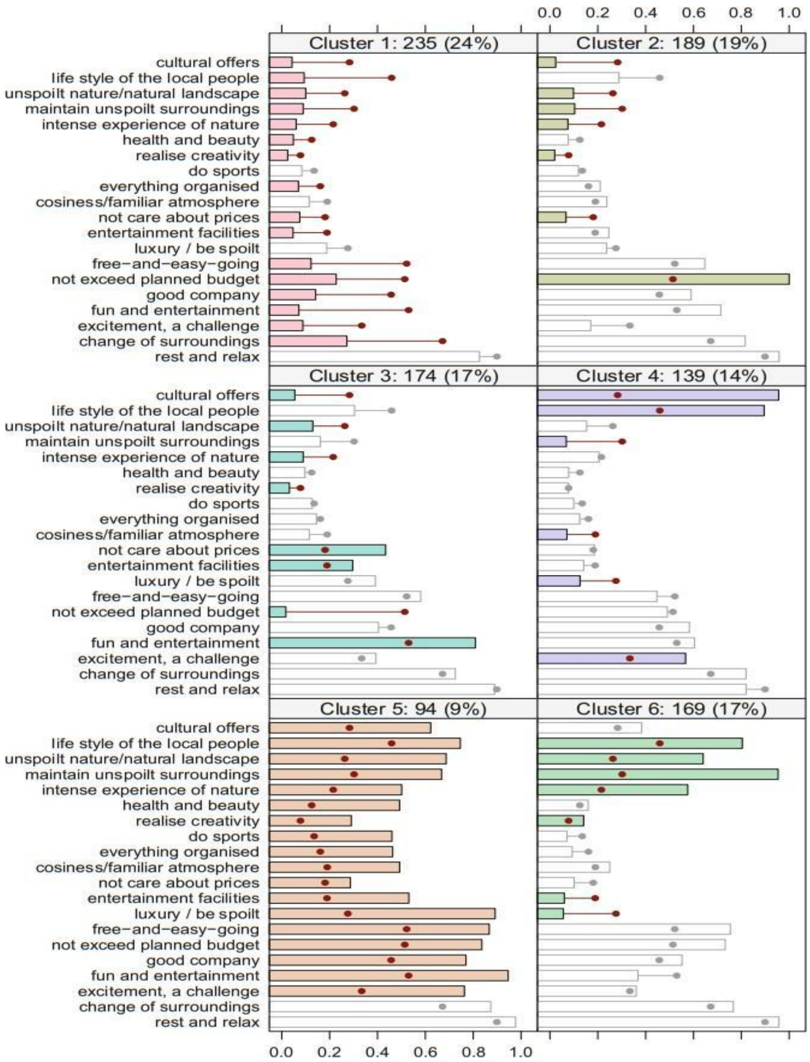


Fig 4 : Segment profile plot for the six-segment solution of the Australian travel motives data set

6.4 Assessing Segment Separation:

A segment separation plot is used to visualize segment separation, depicting the overlap of segments across all relevant dimensions of the data space. While segment separation plots are simple with a low number of segmentation variables, they become complex as the number of variables increases. However, even in complex situations, segment separation plots provide a quick overview of the data and segmentation solution for data analysts and users.

The plots below is showing the cluster of data of some arbitrary dataset.

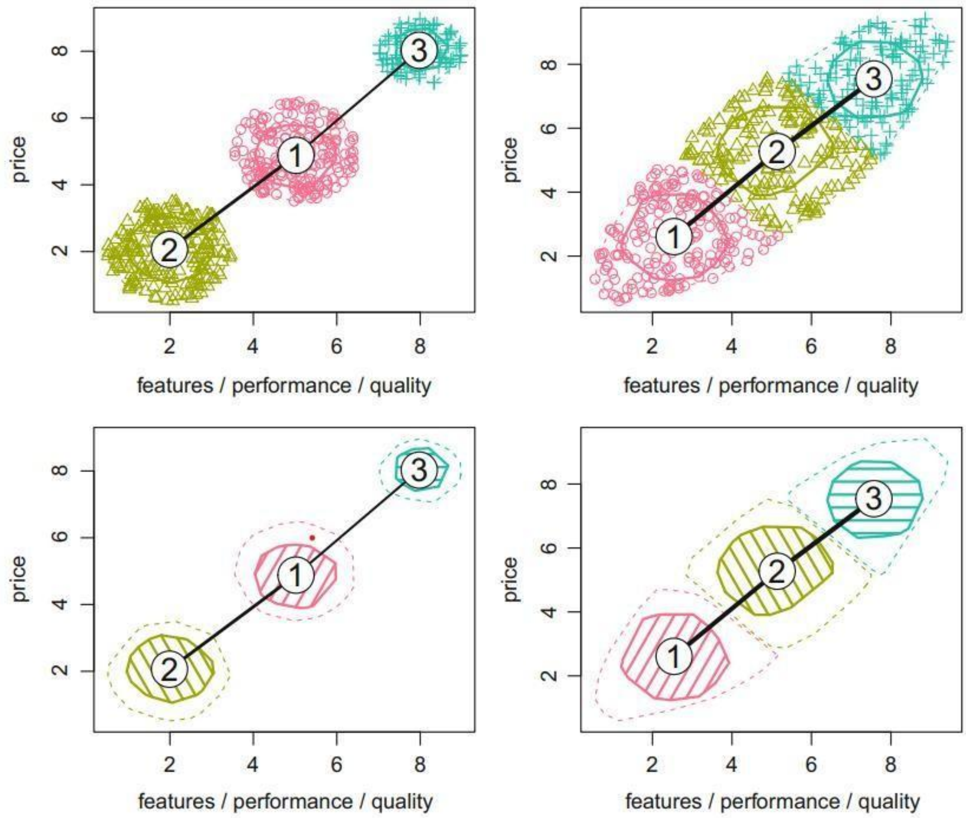


Fig 5 : Segment separation plot including observations (first row) and not including observations (second row) for two artificial data sets: three natural, well-separated clusters (left column); one elliptic cluster (right column)

The excerpt describes the segment separation plot, which is used to visualize the overlap of segments in a data set. The plot consists of a scatter plot of the observations colored by segment membership, along with the projected cluster hulls and a neighborhood graph that indicates the similarity between segments. The plot is demonstrated using two artificial data sets that contain three distinct segments and an elliptic data structure. In the plot, the color of the observations indicates true segment membership, and the different cluster hulls indicate the shape and spread of the true segments. The dashed cluster hulls contain all observations, while the solid cluster hulls contain approximately half of the observations. The neighborhood graphs, represented by black lines with numbered nodes, indicate similarity between segments. The width of the black line is thicker if more observations have these two segment centers as their closest segment centers. The excerpt also notes that in cases where there are high-dimensional data sets, the data may need to be projected onto a smaller number of dimensions to create a segment separation plot using techniques like principal components analysis or separation maximization. The following figure showing some overlapping segments which are usually hard to interpret. But using the PCA we can partially interpret those segment.

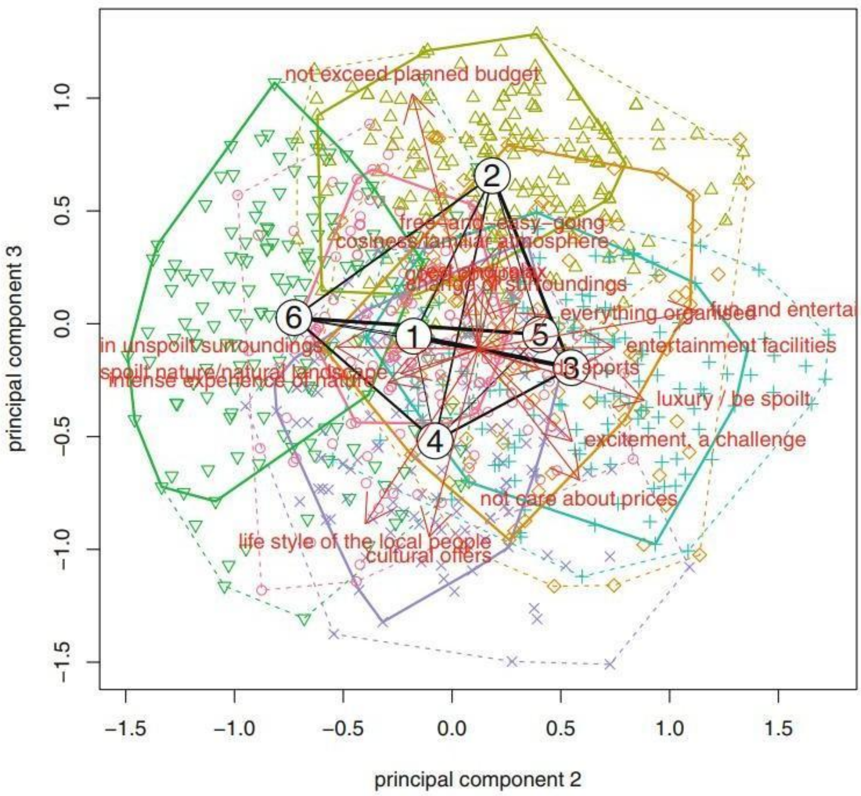


Fig 6 : Segment separation plot using principal components 2 and 3 for the Australian travel motives data set

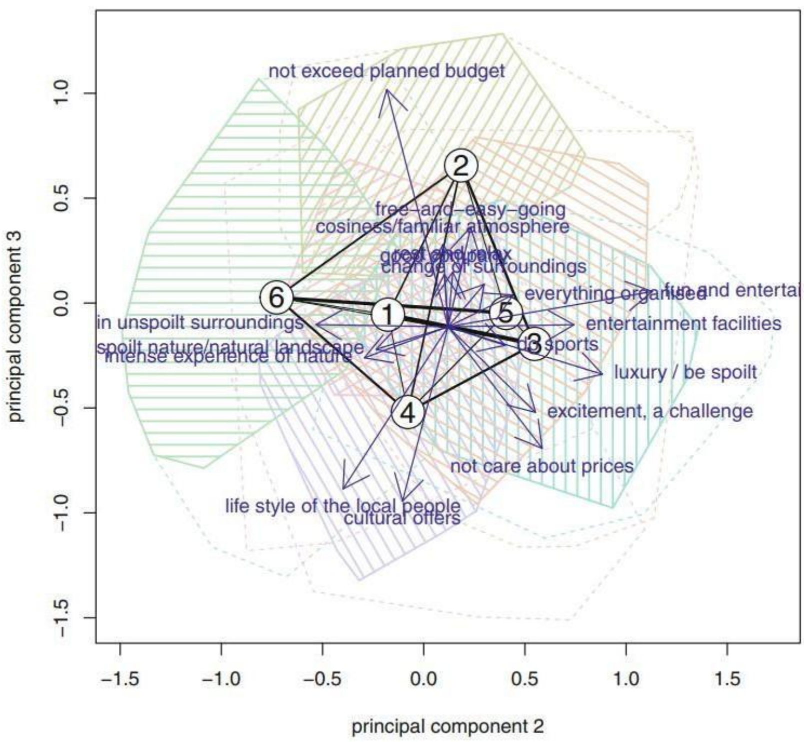


Fig 7 : Segment separation plot using principal components 2 and 3 for the Australian travel motives data set without observations

Each segment separation plot only visualises one possible projection. So, for example, the fact that segments 1 and 5 in this particular projection overlap with other segments does not mean that these segments overlap in all projections. However, the fact that segments 6 and 3 are well-separated in this projection does allow the conclusion – based on this single projection only – that they represent distinctly different tourists in terms of the travel motives.

Step 1: Implications of Committing to Market Segmentation

Implications of Committing to Market Segmentation

The key implication is that the organisation needs to commit to the segmentation strategy for the long term. Market segmentation is a marriage, not a date.

The commitment to market segmentation goes hand in hand with the willingness and ability of the organisation to make substantial changes and investments.

Not to segment unless the expected increase in sales is sufficient to justify implementing a segmentation strategy.

To maximise the benefits of market segmentation – organisations need to organise around market segments, rather than organising around products.

Implementation Barriers

This section highlights the importance of strong leadership and commitment at all levels of an organisation in order to successfully implement a market segmentation strategy. Lack of support from senior leadership can undermine the process and make it difficult for marketing executives to implement the strategy effectively. Other obstacles to successful implementation include resistance to change, lack of creative thinking, poor communication, short-term thinking, and resource limitations. Overall, understanding and commitment to the principles of market segmentation are essential for successful implementation.
The decision to investigate the potential of a market segmentation strategy Must be made at the highest executive level, and must be systematically and continuously communicated and reinforced at all organisational levels and across all organisational units.

Step 2: Specifying the Ideal Target Segment

Segment Evaluation Criteria

This Section provides a list of evaluation criteria that should be considered when conducting market segmentation analysis. The criteria include factors such as segment size, growth potential, competition, profitability, accessibility, compatibility with the company, and measurable differences from other segments. In addition, the involvement of the user at various stages of the process is emphasised. The paragraph also highlights the importance of understanding consumer motivation and goals, assessing financial resources, and considering legal and stakeholder issues. Overall, the criteria aim to ensure that the resulting segments are actionable and capable of developing maximum differential in competitive strategy while preserving competitive advantage.

Knock-Out Criteria

• The segment must be homogeneous; members of the segment must be similar to one another.
• The segment must be distinct; members of the segment must be distinctly different from members of other segments.
• The segment must be large enough; the segment must contain enough consumers to make it worthwhile to spend extra money on customising the marketing mix for them.
• The segment must be matching the strengths of the organisation; the organisation must have the capability to satisfy segment members' needs.
• Members of the segment must be identifiable; it must be possible to spot them in the marketplace.

• The segment must be reachable; there has to be a way to get in touch with members of the segment in order to make the customised marketing mix accessible to them.

Implementing a Structured Process

A large number of possible criteria has to be investigated before agreement is reached on which criteria are most important for the organisation. Using no more than six factors as the basis for calculating these criteria is what's recommended.
the market segmentation team should have a list of
approximately six segment attractiveness criteria. Each of these criteria should have
a weight attached to it to indicate how important it is to the organisation compared
to the other criteria. The typical approach to weigh things is to ask all team members to distribute 100 points across the segmentation criteria. These allocations then have to be negotiated until agreement is reached. Optimally, approval by the advisory committee should

be sought because the advisory committee contains representatives from multiple organisational units bringing a range of different perspectives to the challenge of specifying segment attractiveness criteria.

In summary, there is general agreement in the segmentation literature that a structured process is beneficial when assessing market segments for target market selection. The use of a segment evaluation plot showing segment attractiveness and organisational competitiveness along two axes is a popular approach. However, the criteria used for determining segment attractiveness and organisational competitiveness must be negotiated and agreed upon by the segmentation team, ideally composed of representatives from various organisational units. McDonald and Dunbar recommend using no more than six factors as the basis for calculating these criteria. The team should have a list of approximately six segment attractiveness criteria, each with a weight attached to it, indicating its importance compared to the other criteria. The allocation of weights should be negotiated until agreement is reached. The advisory committee's approval should be sought to ensure that representatives from multiple organisational units provide different perspectives on specifying segment attractiveness criteria.

Step 3: Collecting Data

Segmentation Variables

Table 5.2 Segmentation variables in data-driven market segmentation

Sociodemographics		Travel behaviour	Benefits sought				
gender	age	N° of vacations	relaxation	action	culture	explore	meet people
Female	34	2	1	0	1	0	1
Female	55	3	1	0	1	0	1
Male	87	2	1	0	1	0	1
Female	68	1	0	1	1	0	0
Female	34	1	0	0	1	0	0
Female	22	0	1	0	1	1	1
Female	31	3	1	0	1	1	1
Male	55	4	0	1	0	1	1
Male	43	0	0	1	0	1	0
Male	23	0	0	1	1	0	1
Male	19	3	0	1	1	0	1
Male	64	4	0	0	0	0	0
descriptor variables			segmentation variables				

Table 5.2 shows a data-driven segmentation approach using multiple segmentation variables. In this example, the segmentation variables are gender, age, the number of vacations taken, and information about five benefits people seek or do not seek when they go on vacation. Each row in the table represents one consumer, and the values in the table represent the consumer's responses to the segmentation variables. The data is analysed using statistical techniques to identify segments that share similar characteristics across the segmentation variables.

In this example, three market segments are identified: Segment 1, Segment 2, and Segment 3. Segment 1 is characterised by being predominantly female, older, and taking fewer vacations. They seek relaxation and exploration as benefits of vacation, but not adventure, socialisation, or excitement. Segment 2 is characterised by being predominantly male, middle-aged, and taking many vacations. They seek relaxation, socialisation, and excitement as benefits of vacation, but not exploration or adventure. Segment 3 is characterised by being predominantly female, younger, and taking many vacations. They seek adventure, socialisation, and excitement as benefits of vacation, but not relaxation or exploration.

Descriptor variables are then used to provide a detailed description of each segment. In this example, the descriptor variables include socio-demographic characteristics, as well as information about media behaviour and preferred vacation activities. This information is used to develop a marketing mix that is tailored to the characteristics and needs of each segment.

In summary, empirical data is crucial for both commonsense and data-driven market segmentation. The quality of the data is critical for accurately assigning individuals to market segments and correctly describing the segments. Good market segmentation analysis requires good empirical data, which can come from a variety of sources including surveys, observations such as scanner data, and experimental studies. The optimal source of data for segmentation studies should reflect consumer behaviour, and surveys should not be relied upon solely as they may not always reflect actual behaviour. It is important to explore a range of possible sources to find the best source of data for each segmentation study.

Segmentation Criteria

Choosing the appropriate segmentation criterion depends on the nature of the product or service being offered and the specific marketing context. It requires a thorough understanding of the market and the needs and preferences of consumers. While there is no single "best" criterion to use, it is generally recommended to start with the simplest approach that will effectively identify and target the desired market segments. Demographic and geographic criteria are often used because they are readily available and easy to use, but psychographic and behavioural criteria can provide more nuanced and specific insights into consumer behaviour and preferences. Ultimately, the choice of segmentation criterion should be based on a careful analysis of the market and the available data.

Geographic Segmentation

Geographic segmentation is the oldest and simplest form of market segmentation, where consumers are grouped based on their location of residence. This approach can be useful when language or cultural differences across regions affect consumer behavior, and when communication messages and channels can be easily targeted. However, geographic location alone may not be a sufficient criterion for effective market segmentation, as people within the same region may have different preferences and needs. Despite this limitation, geographic segmentation has been used in recent studies to extract market segments across different countries, which requires careful consideration of cultural differences and biases in data collection.

Socio-Demographic Segmentation

In addition to the limitations mentioned in the previous section, socio-demographic segmentation criteria may also lead to stereotyping and overgeneralization of consumer behavior based on certain demographic characteristics. For example, assuming that all older adults are not interested in technology products, or that all women are interested in cosmetics. This can lead to a lack of understanding of individual preferences and needs, and may result in ineffective marketing strategies.

Furthermore, socio-demographic criteria may not be stable over time, as changes in life circumstances and experiences can result in changes in consumer behavior. For instance, a person's income, education level, and family status may change over time, leading to changes in their product preferences.
Overall, while socio-demographic segmentation can be useful in certain situations, it is important to recognize its limitations and consider other criteria, such as psychographic and behavioral variables, to gain a more comprehensive understanding of consumer behavior.

Psychographic Segmentation

Additionally, psychographic segmentation can sometimes suffer from issues related to self-reported data, such as social desirability bias and response bias. Consumers may not always be honest or accurate in their reporting of their beliefs, interests, and preferences, which can lead to incorrect segmentation decisions. Furthermore, the psychographic approach can be more costly and time-consuming than other segmentation methods because of the need for comprehensive data collection and analysis.
Despite these challenges, psychographic segmentation can provide valuable insights into consumer behavior and preferences. By understanding the underlying motivations and values of different consumer segments, companies can tailor their marketing messages, product offerings, and customer experiences to better meet the needs and desires of their target customers.
Psychographic segmentation can also be particularly effective in industries where emotional and psychological factors play a significant role in consumer decision-making, such as fashion, beauty, and luxury goods.

Behavioural Segmentation

Additionally, behavioural segmentation may not provide insight into the underlying reasons for differences in consumer behaviour, as it only focuses on observable actions. For example, two consumers may purchase a product with similar frequency and spend similar amounts, but have very different motivations for doing so. Therefore, behavioural segmentation may not be the most effective approach for developing marketing strategies that target specific consumer needs and preferences. However, it can be useful in combination with other segmentation approaches, such as psychographic or demographic segmentation.

Choice of Variables

In commonsense segmentation, selecting the appropriate variables is also critical to the quality of the segmentation solution. In this approach, the variables selected should be based on expert knowledge, experience and intuition. Experts in the field should be able to identify the critical dimensions that differentiate customers and the characteristics that are relevant to purchasing behaviour. Once these dimensions have been identified, they can be used as the basis for segmenting the market. The key is to select variables that have a strong theoretical justification, are relevant to the problem at hand, and have a clear practical interpretation. The segmentation solution should also be evaluated against expert judgement to ensure that it is

consistent with prior knowledge and experience.
In summary, selecting the appropriate variables for market segmentation is critical for developing a high-quality solution. In data-driven segmentation, all relevant variables must be included, while unnecessary variables should be avoided to prevent respondent fatigue, reduce dimensionality and prevent masking and noisy variables.
In commonsense segmentation, expert judgement should be used to select relevant variables that have a strong theoretical justification and practical interpretation.

Response Options

The answer options provided to respondents in surveys have an impact on the type of data available for subsequent analyses. Binary or dichotomous responses generate binary data, nominal responses generate nominal data, metric responses generate metric data, and responses with a limited number of ordered answer options generate ordinal data. Binary or metric response options are preferred, as they prevent complications with distance measures in data-driven segmentation analysis. However, if fine nuances of responses need to be captured, visual analogue scales can be used to generate metric data. In many cases, binary response options outperform ordinal answer options, especially when formulated in a level free way.

Response Styles

There are several strategies that can be employed to minimize the risk of capturing response styles in survey data for market segmentation. First, it is important to use clear and unambiguous question wording to avoid confusion or misinterpretation by respondents. Second, including reverse-coded items (i.e., items with the opposite polarity) can help identify respondents who are using response styles, such as extreme or midpoint responding, rather than providing thoughtful answers. Third, using different question formats, such as open-ended questions or ranking tasks, can provide additional information and reduce reliance on rating scales. Fourth, pilot testing the survey with a sample of the target population can help identify and address response biases and issues with question wording. Finally, data cleaning and analysis techniques, such as identifying and removing respondents who consistently display a response style, can help ensure that market segmentation is based on meaningful differences in respondent characteristics and behaviors, rather than artifacts of the survey design or response biases.

Sample Size

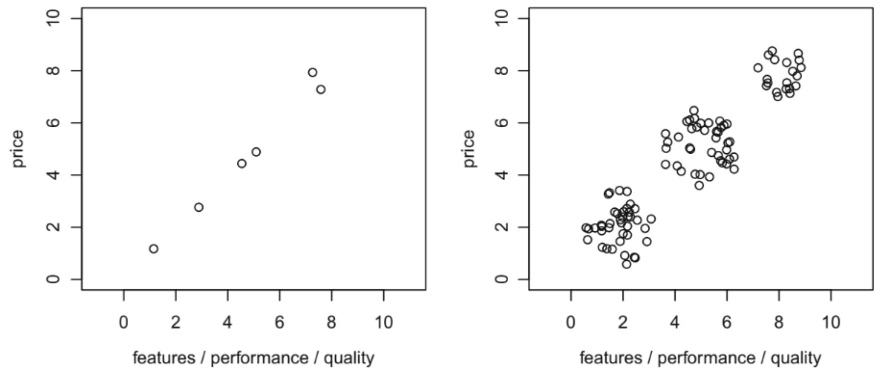


Fig. 5.1 Illustrating the importance of sufficient sample size in market segmentation analysis

There is no one-size-fits-all sample size recommendation for market segmentation analysis. The appropriate sample size depends on various factors, including the number of segmentation variables, the number and size of segments, and the specific algorithm and inference method used. However, some researchers have proposed rules of thumb, such as having at least 2p or five times 2p observations per segment for latent class analysis using binary variables, or having a sample size of at least 10 · p · k for constructing artificial data sets to study clustering algorithms, with a minimum sample size of 10 · p for the smallest segment in unequally sized segments. Ultimately, it is important to use a sample size that is sufficient to detect meaningful differences between segments while avoiding capturing random variation or response styles.

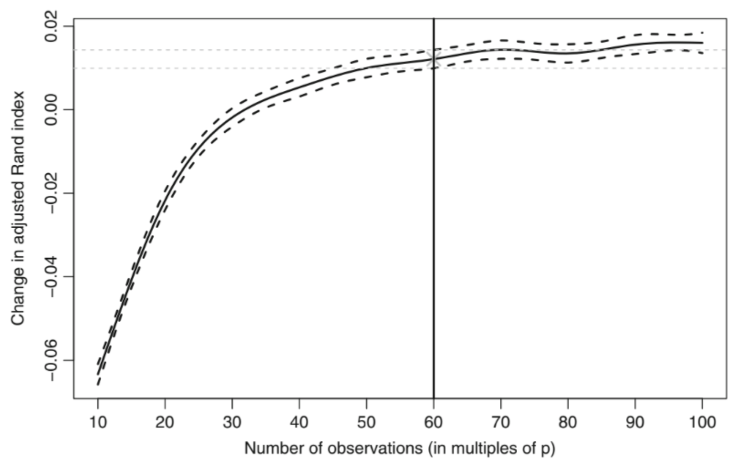


Fig. 5.2 Effect of sample size on the correctness of segment recovery in artificial data. (Modified from Dolnicar et al. 2014)

Figure 5.2 shows that the correctness of segment recovery increases with an increase in sample size. For example, for a data set with four segments and two segmentation variables, a sample size of at least 50 times the number of segmentation variables (i.e., 100 in this case) is needed to achieve an adjusted Rand index of 0.8 or higher. In contrast, a sample size of only 10 times the number of segmentation variables (i.e., 20 in this case) results in an adjusted Rand index of only 0.3, indicating poor alignment between the true segment solution and the extracted one. These findings highlight the importance of sample size in accurately identifying market segments and suggest that a larger sample size is generally better for this task.
Data characteristics studied included: the number of segmentation variables, the correlation among segmentation variables, and the scale used to measure the segmentation variables. The results of the simulation study by Dolnicar et al. (2016)

showed that sample size requirements increase substantially when market segments overlap or are unequally sized. In the presence of both features, the sample size required for accurate segmentation solutions is very high.
Furthermore, the study found that sample size requirements increase with the number of segmentation variables, the level of correlation among segmentation variables, and the use of ordinal or nominal scales to measure segmentation variables. Specifically, the study recommends a sample size of at least 100 · p for segmentation studies using nominal scales, and at least 150 · p for studies using ordinal scales.

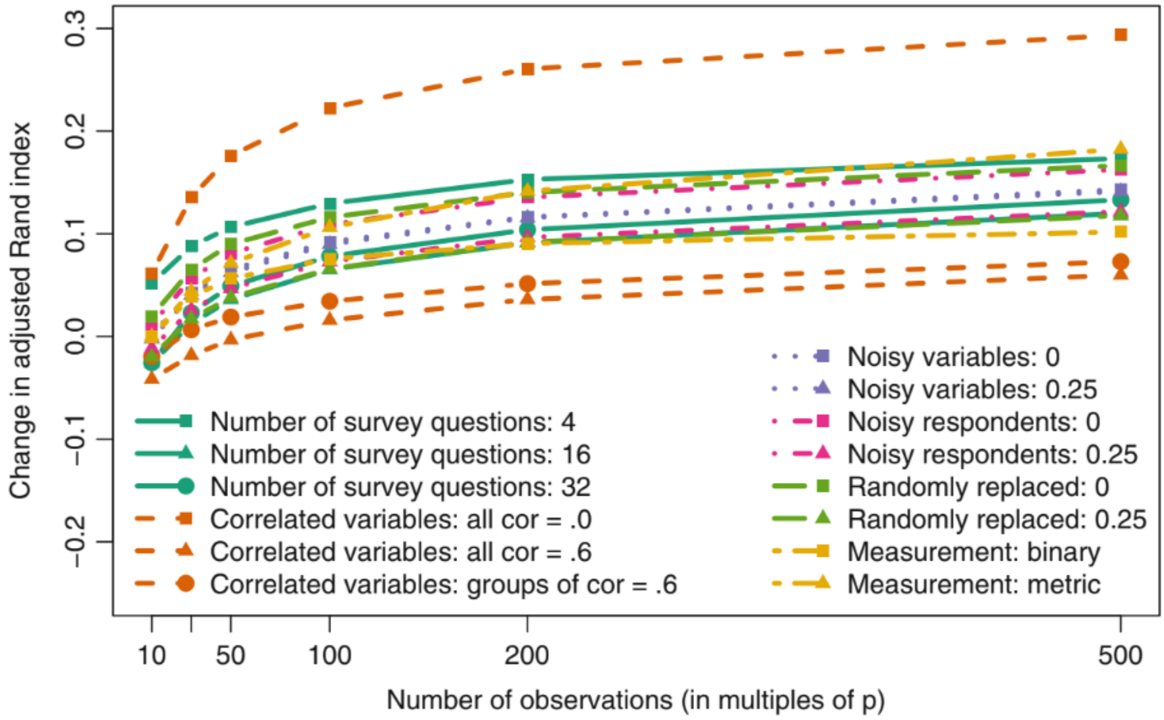


Fig. 5.3 shows that increasing sample size improves the ability of segmentation algorithms to correctly identify market segments, but the extent to which this is the case varies depending on market and data characteristics. Some characteristics, such as uncorrelated segmentation variables, can be compensated for by increasing sample size, while others, such as high correlation between variables, cannot. Additionally, characteristics such as sampling error, response biases and styles, and low data quality can also affect segment recovery, and may require additional attention beyond increasing sample size to address.

the quality of market segmentation results based on such data that, optimally, data used in market segmentation analyses should

- contain all necessary items;
- contain no unnecessary items;
- contain no correlated items;
- contain high-quality responses;
- be binary or metric;
- be free of response styles;
- include responses from a suitable sample given the aim of the segmentation study; and
- include a sufficient sample size given the number of segmentation variables (100 times the number of segmentation variables).

Data from Internal Sources

Additionally, internal data may not capture certain aspects of consumer behavior that are important for segmentation analysis, such as psychographic or attitudinal variables. This can limit the ability of segmentation models based on internal data to capture the full range of consumer behavior and preferences.
Moreover, internal data may not provide a representative sample of the market as a whole, since it only includes data from consumers who have interacted with the organization in some way. This can result in biased or incomplete segmentations that may not generalize well to the broader market. It is important, therefore, to supplement internal data with external data sources to ensure a more representative and complete picture of the market.
Despite these challenges, internal data can still be a valuable resource for market segmentation analysis, particularly when combined with external data sources. By leveraging internal data, organizations can gain valuable insights into the behavior and preferences of their existing customers, and use this information to develop more targeted marketing strategies and improve customer retention.

Data from Experimental Studies

Experimental data can be a valuable source of information for market segmentation analysis as they provide insights into consumer behavior and preferences that may not be captured through surveys or internal data. By manipulating specific stimuli or attributes, experiments can help identify the key drivers of consumer behavior and preferences. Conjoint analysis, for example, is a powerful tool that can be used to determine the relative importance of different attributes and attribute levels in consumer decision making.

Step 7: Describing Segments

9.1 Developing a Complete Picture of Market Segments

Segment profiling involves understanding differences in segmentation variables across market segments. These variables are chosen early in the market segmentation analysis process during the specification of the ideal target segment and the collection of data. Segmentation variables are used to extract market segments from the empirical data.
Step 7 of the market segmentation process involves describing segments. This step is similar to profiling, but it involves the use of additional information to describe market segments using variables other than the segmentation variables used to extract the segments. This step involves crossing segment variables with psychographic, demographic, socio-economic variables, media exposure, and specific product and brand attitudes or evaluations.

For example, in the Australian travel motives data set, profiling involves investigating differences between segments with respect to travel motives. Segment description uses additional information such as age, gender, past travel behavior, preferred vacation activities, media use, use of information sources during vacation planning, or expenditure patterns during a vacation. These additional variables are known as descriptor variables.

Descriptive statistics and visualisations are two methods for studying differences between market segments with respect to descriptor variables. Descriptive statistics provide numerical summaries of the data, such as mean, median, and standard deviation, which can be used to compare segments. Tables and graphs can be used to present the results of descriptive statistics. On the other hand, visualisations provide an easy-to-understand way to convey complex information. For example, bar charts, pie charts, and scatterplots can be used to compare segment characteristics and identify patterns in the data. Visualisations can help marketers quickly identify key differences between segments and develop targeted marketing strategies accordingly.

Using Visualisations to Describe Market Segments

Cornelius et al. (2010) have argued that graphical representations serve to transmit the very essence of marketing research results. They also found that managers prefer graphical formats and view the intuitiveness of graphical displays as critically important. Section 8.3.1 provides an illustration of the higher efficiency with which people process graphical as opposed to tabular results. Overall, graphical statistics are a powerful tool for describing market segments and communicating research results to a wider audience

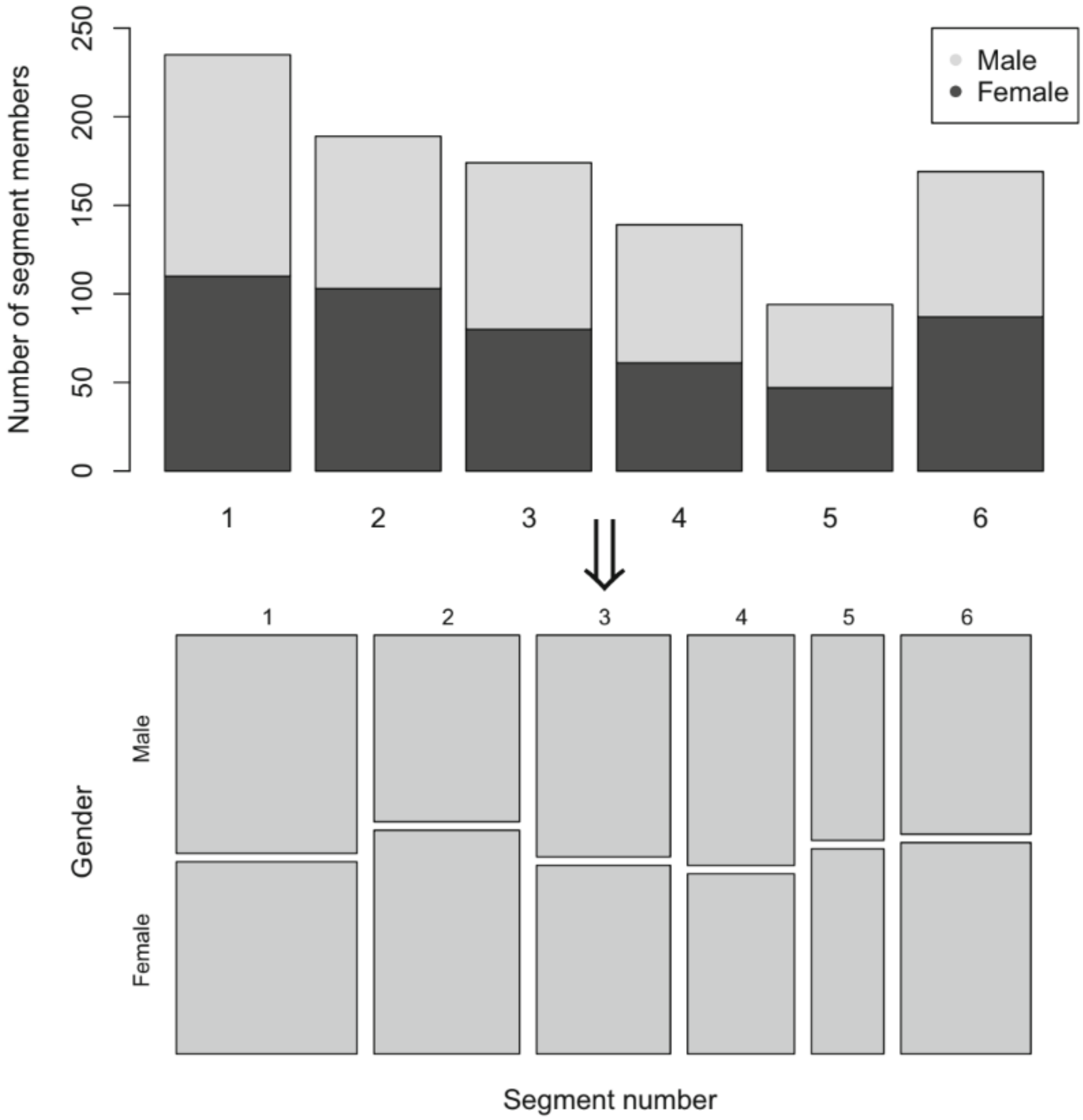
Nominal and Ordinal Descriptor Variables

When describing differences between market segments in one single nominal or ordinal descriptor variable, the basis for all visualisations and statistical tests is a cross-tabulation of segment membership with the descriptor variable.
The easiest approach to generating a cross-tabulation is to add segment membership as a categorical variable to the data frame of descriptor variables. Then we can use the formula interface of R for testing or plotting:
R> vacmotdesc$C6 <- as.factor(C6)

The following R command gives the number of females and males across market segments:

```
R> C6.Gender <- with(vacmotdesc,
+ table("Segment number" = C6, Gender))
R> C6.Gender
Gender
```

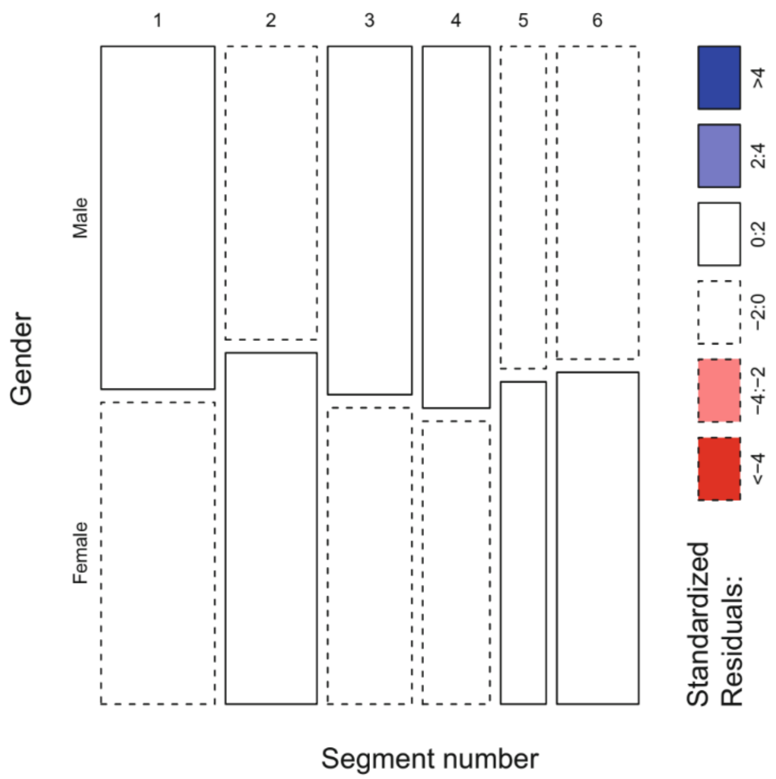
Segment number	Male	Female
1	125	110
2	86	103
3	94	80
4	78	61
5	47	47
6	82	87



By default, function mosaicplot() in R uses dark red cell colouring for contributions or standardised Pearson residuals smaller than −4, light red if contributions are smaller than −2, white (not interesting) between −2 and 2, light blue if contributions are larger than 2, and dark blue if they are larger than 4. Figure 9.2 shows such a plot with the colour coding included in the legend.
In Fig. 9.2 all cells are white, indicating that the six market segments extracted from the Australian travel motives data set do not significantly differ in gender distribution. The proportion of female and male tourists is approximately the same across segments. The dashed and solid borders of the rectangles indicate that the number of respondents in those cells are either lower than expected (dashed

borders), or higher than expected (solid black borders). But, irrespective of the borders, white rectangles mean differences are statistically insignificant.

Figure 9.3 shows that segment membership and income are moderately associated. The top row corresponds to the lowest income category (less than AUD



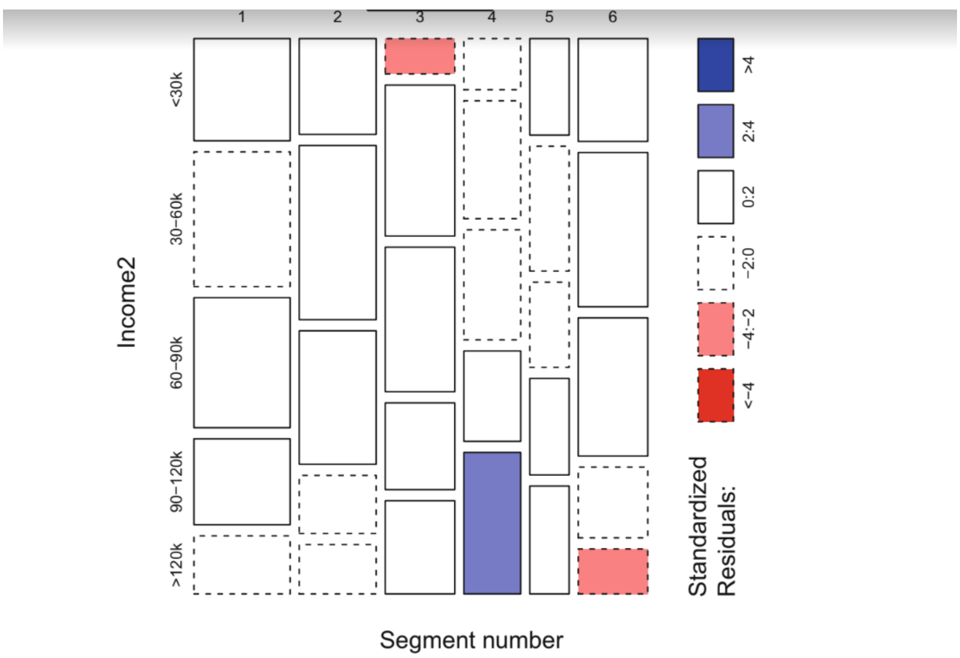


Fig. 9.3 Shaded mosaic plot for cross-tabulation of segment membership and income for the Australian travel motives data set

30,000 per annum). The bottom row corresponds to the highest income category (more than AUD 120,000 per annum). The remaining three categories represent AUD 30,000 brackets in-between those two extremes. We learn that members of segment 4 (column 4 in Fig. 9.3) – those motivated by cultural offers and interested in local people – earn more money. Low income tourists (top row of Fig. 9.3) are less frequently members of market segment 3, those who do not care about prices and instead seek luxury, fun and entertainment, and wish to be spoilt when on vacation. Segment 6 (column 6 in Fig. 9.3) – the nature loving segment – contains fewer members on very high incomes.

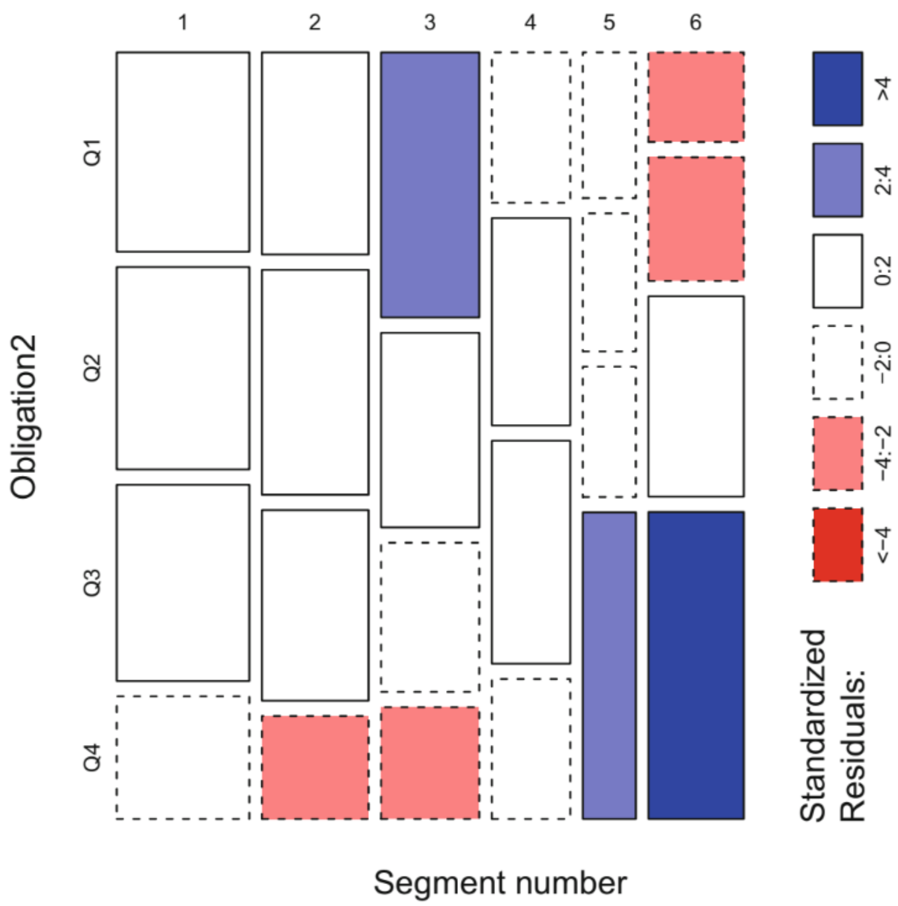


Fig. 9.4 Shaded mosaic plot for cross-tabulation of segment membership and moral obligation to protect the environment for the Australian travel motives data set

Figure 9.4 graphically illustrates the cross-tabulation, associating segment membership and stated moral obligation to protect the environment in a mosaic plot.

Segment 3 (column 3 of Fig. 9.4) – whose members seek entertainment – contains significantly more members with low stated moral obligation to behave in an environmentally friendly way. Segment 3 also contains significantly fewer members in the high moral obligation category. The exact opposite applies to segment 6. Members of this segment are motivated by nature, and plotted in column 6 of Fig. 9.4. Being a member of segment 6 implies a positive association with high moral obligation to behave environmentally friendly, and a negative association with membership in the lowest moral obligation category.

Metric Descriptor Variables

To have segment names (rather than only segment numbers) displayed in the plot, we create a new factor variable by pasting together the word "Segment" and the segment numbers from C6. We then generate a histogram for age for each segment. Argument as.table controls whether the panels are included by starting on the top left (TRUE) or bottom left (FALSE, the default).

```
R> library("lattice")
R> histogram(~ Age | factor(paste("Segment", C6)),
+ data = vacmotdesc, as.table = TRUE)
```

We do the same for moral obligation:

```
R> histogram(~ Obligation | factor(paste("Segment",C6)),
+ data = vacmotdesc, as.table = TRUE)
```

The resulting histograms are shown in Figs. 9.5 (for age) and 9.6 (for moral obligation). In both cases, the differences between market segments are difficult to assess just by looking at the plots.

We can gain additional insights by using a parallel box-and-whisker plot; it shows the distribution of the variable separately for each segment. We create this parallel box-and-whisker plot for age by market segment in R with the following command:

```
R> boxplot(Age ~ C6, data = vacmotdesc,
+ xlab = "Segment number", ylab = "Age")
```

where arguments xlab and ylab customise the axis labels.

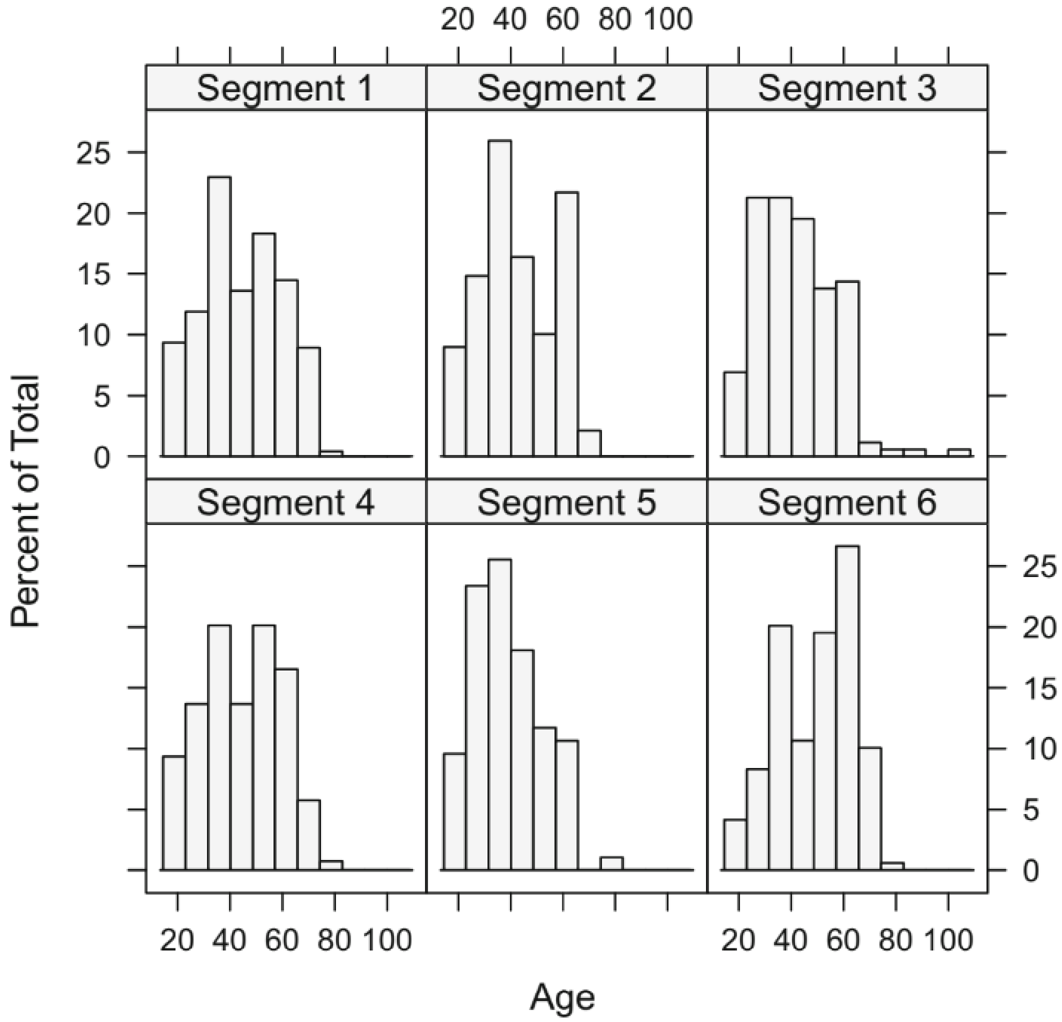
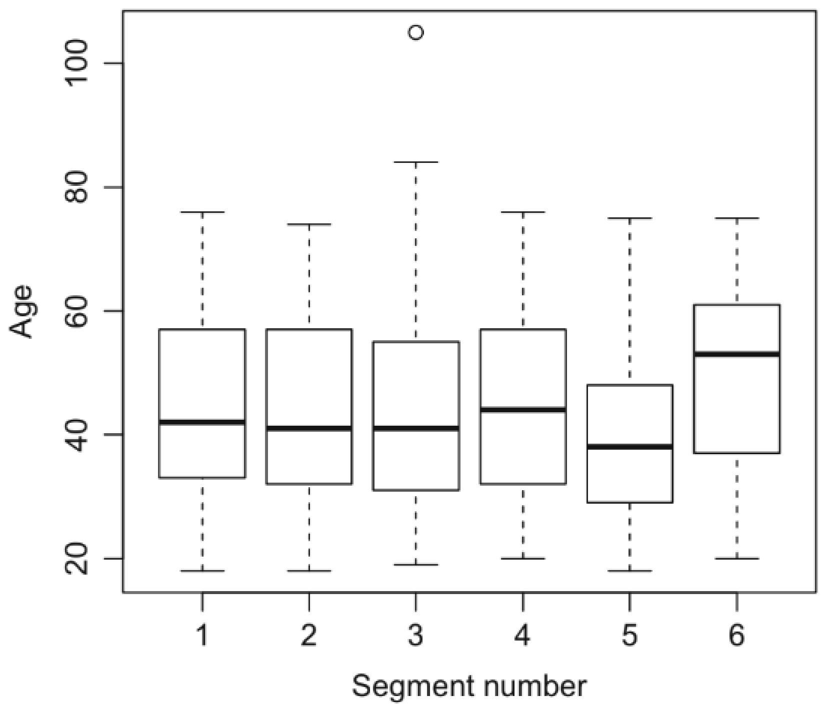


Fig. 9.5 Histograms of age by segment for the Australian travel motives data set

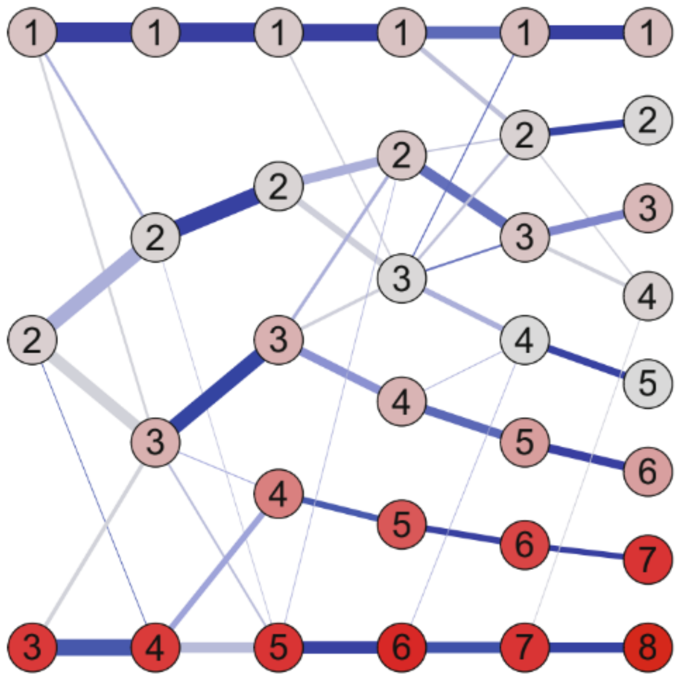
The notches in this version of the parallel box-and-whisker plot correspond to 95% confidence intervals for the medians. If the notches for different segments do not overlap, a formal statistical test will usually result in a significant difference. We can conclude from the inspection of the plot in Fig. 9.8 alone, therefore, that there is a significant difference in moral obligation to protect the environment between members of segment 3 and members of segment 6.

Fig. 9.7 Parallel box-and-whisker plot of age by segment for the Australian travel motives data set



The modified segment level stability across solutions (SLSA) plot is used to represent the stability of segments across multiple solutions. In this plot, the node colour has a different meaning, and the shading of the edges represents the numeric SLSA value. Low stability values are represented by light grey edges, while high stability values are represented by dark blue edges.

Fig. 9.9 Segment level stability across solutions (SLS_A) plot for the Australian travel motives data set for three to eight segments with nodes coloured by mean moral obligation values



Testing for Segment Differences in Descriptor Variables

The χ2-test is a statistical test that is used to determine whether there is a significant association between two categorical variables. It is commonly used to test for independence between rows and columns of a contingency table. The test calculates the difference between the observed frequencies and the expected frequencies and measures how likely it is to observe such a difference by chance.

To perform the χ2-test, the researcher first constructs a contingency table that shows the distribution of the two categorical variables. Then, the expected frequencies are calculated assuming that the two variables are independent. The expected frequency for each cell of the contingency table is calculated as the product of the row total and the column total divided by the grand total. The test statistic is then calculated as the sum of the squared differences between the observed and expected frequencies divided by the expected frequencies.

If the calculated test statistic is larger than the critical value of the χ2-distribution at a given level of significance and degrees of freedom, the researcher can reject the null hypothesis that there is no association between the two variables and conclude that there is a significant association between the variables.

The p-value indicates how likely the observed frequencies occur if there is no association between the two variables (and sample size, segment sizes, and overall gender distribution are fixed). Small p-values (typically smaller than 0.05), are taken as statistical evidence of differences in the gender distribution between segments. Here, this test results in a non-significant p-value, implying that the null hypothesis is not rejected. The mosaic plot in Fig. 9.2 confirms this: no effects are visible and no cells are coloured.

If the χ2-test rejects the null hypothesis of independence because the p-value is smaller than 0.05, a mosaic plot is the easiest way of identifying the reason for rejection. The colour of the cells points to combinations occurring more or less frequently than expected under independence.

The association between segment membership and metric variables (such as age, number of nights at the tourist destinations, dollars spent on accommodation) is visualised using parallel boxplots. Any test for difference between the location (mean, median) of multiple market segments can assess if the observed differences in location are statistically significant.

The most popular method for testing for significant differences in the means of more than two groups is Analysis of Variance (ANOVA).

After rejecting the null hypothesis of the analysis of variance, we need to conduct pairwise comparisons between segments to identify which ones have significantly different means for the variable of interest (in this case, moral obligation to protect the environment). These comparisons can be done using post-hoc tests, such as Tukey's HSD (Honestly Significant Difference) test or Bonferroni correction. These tests adjust the p-values for multiple comparisons and provide a more accurate assessment of which segments are different from each other.

Multiple testing occurs when multiple comparisons are made on the same data set. It increases the probability of obtaining at least one false positive result by chance alone, even if all the individual tests are done correctly. Therefore, it is necessary to adjust the p-values to control for the overall false positive rate when multiple comparisons are made.

There are different methods available to adjust the p-values for multiple testing, such as the Bonferroni correction and the Holm-Bonferroni method. The Bonferroni correction is a very conservative approach that multiplies all p-values by the number of tests computed. This method is too stringent in most cases, as it reduces the probability of rejecting the null hypothesis too much.

The Holm-Bonferroni method is a less conservative method that adjusts the p-values by ranking them in ascending order and multiplying them by the number of remaining comparisons.
Another commonly used method is the false discovery rate (FDR) procedure, proposed by Benjamini and Hochberg (1995). This method controls the expected proportion of false positives among the total number of rejections and is less conservative than the Bonferroni correction.
In R, the p.adjust() function can be used to adjust the p-values for multiple testing, and it offers different methods to do so, including the Bonferroni and Holm-Bonferroni methods and the FDR procedure. It is important to adjust the p-values for multiple testing to avoid false positive results and to increase the reliability of the statistical inference.

The Tukey HSD (honestly significant difference) test is used to compare all pairs of means in an ANOVA analysis. The R code plot(TukeyHSD(aov1), las = 1) produces a plot of the results of the Tukey HSD test for the market segmentation example.
The resulting plot (Figure 9.10) shows the point estimate of the difference in mean values for each pairwise comparison in the middle of a horizontal solid line. The length of the line represents the confidence interval for the difference in means, adjusted for the multiple comparisons being made. If the confidence interval crosses the vertical line at 0, the difference is not significant. If the confidence interval does not cross the vertical line at 0, the difference is significant.
In the market segmentation example, the plot shows that segments 1, 2, 3, and 4 do not differ significantly in moral obligation, and segments 5 and 6 have a significantly higher moral obligation than the other segments (except for the non-significant difference between segments 4 and 5). Segment 4 sits between the low and high moral obligation groups and does not differ significantly from segments 1-3 at the low end and segment 5 at the high end of the moral obligation range.
Overall, the Tukey HSD test confirms the results of the pairwise t-tests and provides additional information on the significance of the differences in means between all pairs of segments.

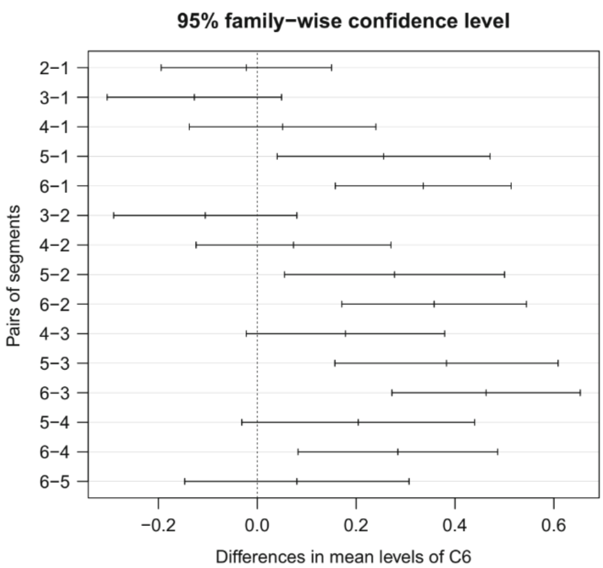


Fig. 9.10 Tukey's honest significant differences of moral obligation to behave environmentally friendly between the six se[redacted]a set

Predicting Segments from Descriptor Variables

It is not accurate to say that including the intercept β0 in the model formula drops the regression coefficient for segment 1. The intercept represents the average value of the dependent variable when all independent variables are equal to zero (or their reference levels, in the case of categorical variables). In the case of a categorical independent variable with k categories, the intercept represents the mean value of the dependent variable for the reference category (i.e., the category that is not explicitly included in the formula).

Therefore, including the intercept in the model formula does not drop any regression coefficient. Instead, it estimates the average value of the dependent variable for the reference category, and estimates the difference between the means of the dependent variable for each of the other categories and the reference category. In other words, it estimates the effect of each category relative to the reference category.
The passage discusses the use of regression models in statistical analysis. Linear regression models are used to estimate the relationship between a dependent variable and one or more independent variables. The regression coefficients indicate how much the dependent variable changes when one independent variable changes, assuming all other independent variables remain constant. The linear regression model assumes that changes in one independent variable are independent of the absolute level of all independent variables. Generalized linear models are used to accommodate a wider range of distributions for the dependent variable, particularly when the dependent variable is categorical and the normal distribution is not appropriate.

Binary Logistic Regression

The logit link can be used to map the success probability of binary data onto $(-\infty, \infty)$ in generalized linear models. In R, the glm() function is used to fit generalized linear models, with the family argument specifying the distribution of the dependent variable and the link function. The binomial distribution with logit link is used for binary data, and the model is specified using the formula interface with the dependent variable on the left of ~ and the independent variables on the right. The I() function is used to construct a binary indicator variable for the dependent variable.
The plot on the right side of Figure 9.11 illustrates that the probability of being a member of segment 3 decreases as the moral obligation increases. Based on the plot, respondents of average age with a moral obligation value of Q1 have a predicted probability of about 25% of belonging to segment 3. If these respondents have the highest moral obligation value of Q4, their predicted probability decreases to 12%. The 95% confidence intervals of the estimated effects reveal that despite high uncertainty, probabilities do not overlap for the two most extreme values of moral obligation. This indicates that incorporating moral obligation into the logistic regression model enhances model fit.

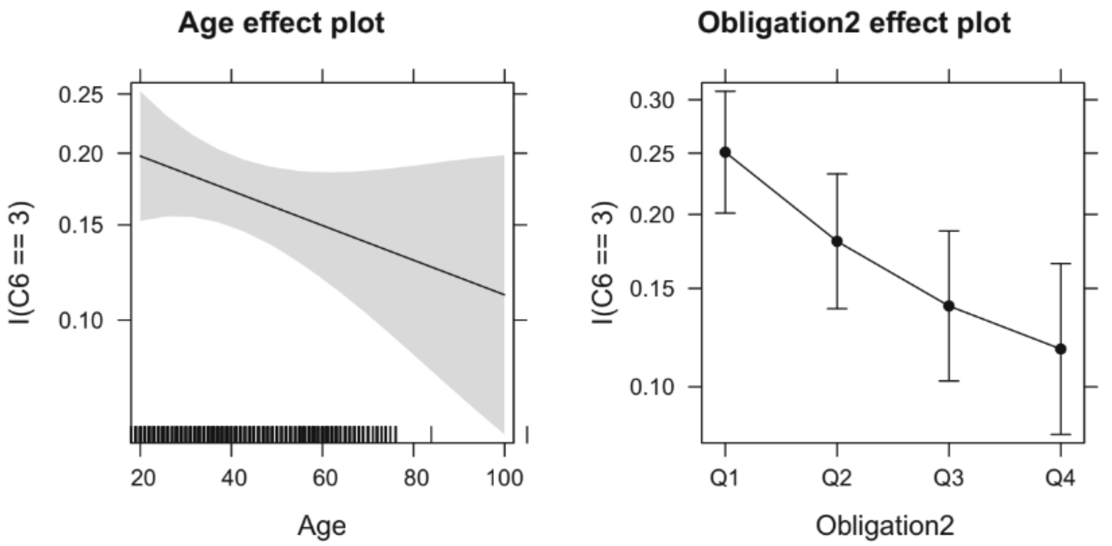


Fig. 9.11 Effect visualisation of age and moral obligation for predicting segment 3 using binary logistic regression for the Australian travel motives data set

Multinomial Logistic Regression

Multinomial logistic regression is a statistical model used to predict the probability of belonging to one of several nominal categories (more than two categories) based on one or more independent variables. It is an extension of binary logistic regression, which predicts the probability of belonging to one of two categories. In multinomial logistic regression, the dependent variable is a categorical variable with more than two categories, and the independent variables can be continuous or categorical. The model estimates the probability of each category relative to a reference category, based on the values of the independent variables. The multinomial logistic regression model uses a set of coefficients to estimate the effect of each independent variable on the probability of belonging to each category. The coefficients are estimated using maximum likelihood estimation, and the model can be evaluated using measures of goodness of fit such as the deviance or the likelihood ratio test. Multinomial logistic regression is commonly used in social science, health, and marketing research to analyze data where the dependent variable is categorical with more than two categories.
With function Anova() we assess if dropping a single variable significantly

reduces model fit. Dropping a variable corresponds to setting all regression coef-
ficients of this variable to 0. This means that the regression coefficients in one or

several columns of the regression coefficient matrix corresponding to this variable
are set to 0. Function Anova() tests if dropping any of the variables significantly
reduces model fit.
Starting with the full model containing all available independent variables, the stepwise procedure returns the best-fitting model, the model which deteriorates
in AIC if an independent variable is either dropped or additionally included.
We assess the predictive performance of the fitted model by comparing the
predicted segment membership to the observed segment membership. Figure 9.13

shows a mosaic plot of the predicted and observed segment memberships on the left. In addition, we investigate the distribution of the predicted probabilities for each Segment.

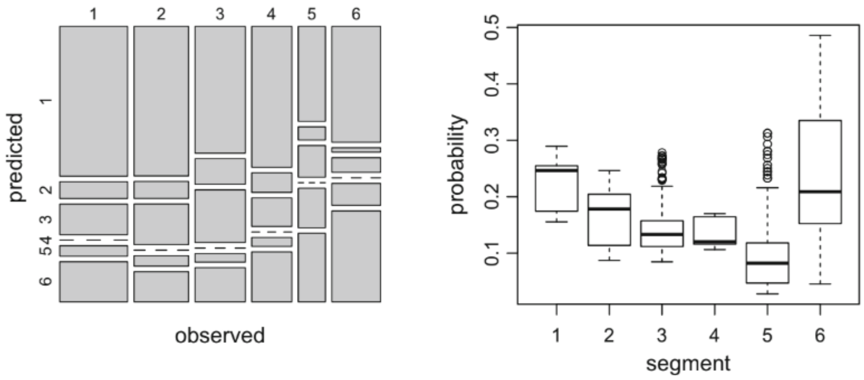


Fig. 9.13 Assessment of predictive performance of the multinomial logistic regression model including age and moral obligation as independent variables for the Australian travel motives data set. The mosaic plot of the cross-tabulation of observed and predicted segment memberships is on the left. The parallel boxplot of the predicted probabilities by segment for consumers assigned to segment 6 is on the right

Figure 9.13 shows parallel boxplots of the predicted segment probabilities for consumers assigned to segment 6 on the right:

```
R> par(mfrow = c(1, 2))
R> pred.class.C6 <-
predict(model.C6)
R> plot(table(observed
= vacmotdesc$C6,
+ predicted = pred.class.C6), main = "")
R> pred.prob.C6 <- predict(model.C6, type = "prob")
R> predicted <- data.frame(prob = as.vector(pred.prob.C6),
+ observed = C6,
+ predicted = rep(1:6, each = length(C6)))
R> boxplot(prob ~ predicted,
+ xlab = "segment", ylab = "probability",
+ data = subset(predicted, observed == 6))
```

The right panel in Fig. 9.14 shows how the predicted segment membership probability changes with moral obligation values for a consumer of average age. The predicted probability to belong to segment 6 increases with increasing moral obligation value. Respondents with the lowest moral obligation value of Q1 have a probability of about 8% to be from segment 6. This increases to 29% for respondents with a moral obligation value of Q4. For segment 3 the reverse is true: respondents with higher moral obligation values have lower probabilities to be from segment 3.

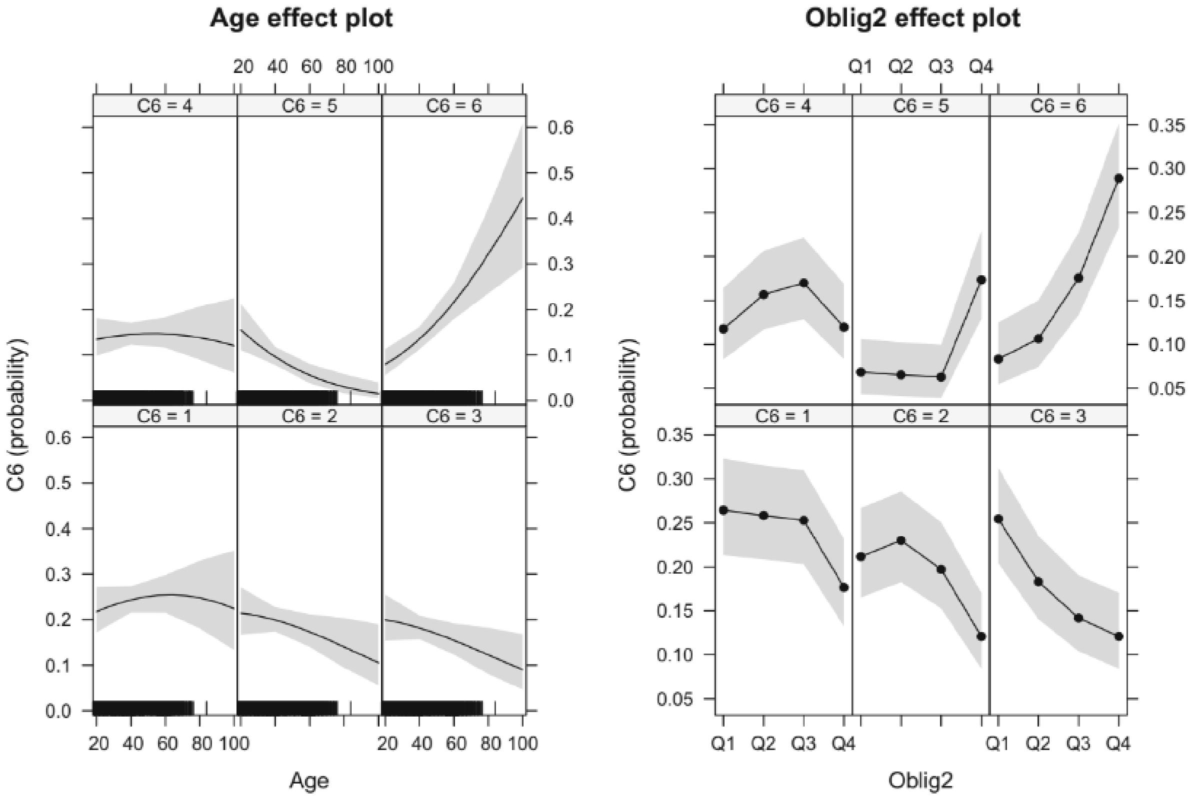


Fig. 9.14 Effect visualisation of age and moral obligation for predicting segment membership using multinomial logistic regression for the Australian travel motives data set

Tree-Based Methods

Classification and regression trees (CARTs) are a machine learning technique used for predicting a binary or categorical dependent variable based on a set of independent variables. The CART approach involves splitting consumers into groups based on one independent variable at each step, with the aim of making the resulting groups as pure as possible with respect to the dependent variable. The resulting tree shows the nodes that emerge from each splitting step, with the root node containing all consumers and terminal nodes containing groups of consumers with similar values for the dependent variable. Segment membership can be predicted by moving down the tree, following the branch that reflects the consumer's independent variable values.
There are several tree constructing algorithms that differ in terms of binary vs. multi-way splits, selection criteria for the independent variable and split point, stopping criteria for the stepwise procedure, and final prediction at the terminal node. The rpart package implements the original CART algorithm proposed by Breiman et al., while the partykit package offers an alternative tree constructing procedure that performs unbiased variable selection based on association tests and p-values. The partykit package also allows for visualisation of fitted tree models, and the ctree() function can be used to fit a conditional inference tree.

The classification and regression tree (CART) approach for predicting a binary or categorical dependent variable based on a set of independent variables. CART models are a supervised learning technique in machine learning that work well with a large number of independent variables, offer ease of interpretation supported by visualizations, and incorporate interaction effects. The approach uses a stepwise procedure to fit the model by splitting consumers into groups based on one independent variable, with the aim of resulting groups being as pure as possible with respect to the dependent variable. The resulting tree shows the nodes that emerge from each splitting step, with nodes that are not split further being terminal nodes. The passage also discusses the differences in tree constructing algorithms, such as splits into two or more groups at each node, selection criteria for independent variables and split points, and stopping criteria for the stepwise procedure. The passage also shows examples of R packages that implement tree constructing algorithms, such as rpart and partykit, with function ctree() from package partykit fitting a conditional inference tree. The passage concludes by discussing the visual representation of the classification tree and the parameters that influence tree construction.

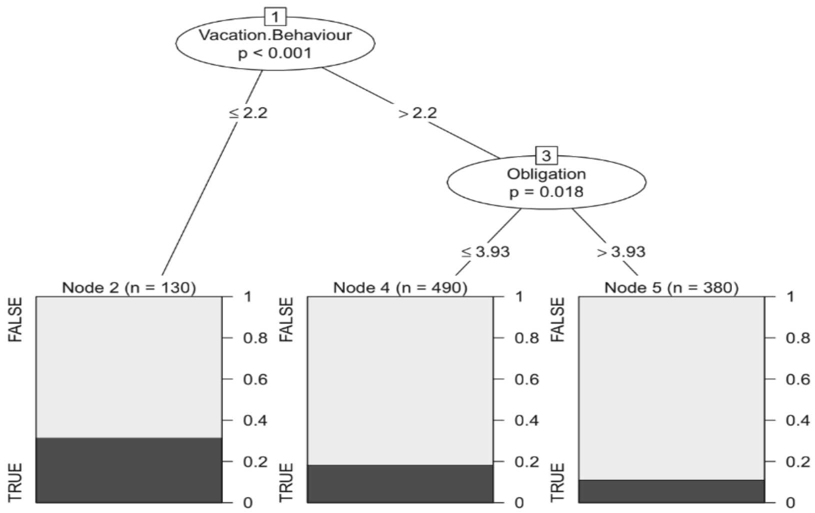


Fig. 9.15 Conditional inference tree using membership in segment 3 as dependent variable for the Australian travel motives data set

Marketing planning is a logical sequence and a series of activities leading to the setting of marketing objectives and the formulation of plans to achieving them. A marketing plan consists of two components: a strategic and a tactical marketing plan. The strategic plan outlines the long-term direction of an organisation, but does not provide much detail on shortterm marketing action required to move in this long-term direction. The tactical marketing plan does the opposite. It translates the long-term strategic plan into detailed instructions for short-term marketing action. The strategic marketing plan states where the organisation wants to go and why. The tactical marketing plan contains instructions on what needs to be done to get there.

Before starting a hike, it is critically important to organise a map, and figure out where exactly one's present location is. Once the present location is known, the next step is to decide which mountain to climb. The choice of the mountain is a strategic decision; it determines all subsequent decisions. As soon as this strategic decision is made, the expedition team can move on to tactical decisions, such as: which shoes to wear for this particular hike, which time of day to depart, and how much food and drink to pack. All these tactical decisions are important to ensure a safe expedition, but they depend entirely on the strategic decision of which mountain to climb.

The tactical marketing plan depends entirely on the strategic marketing plan, but the strategic marketing plan does not depend on the tactical marketing plan.

Approaches to Market Segmentation Analysis

1. Based on Organisational Constraints
2. Based on the Choice of the Segmentation Variables

STEP 1: Deciding (not) to Segment

Implications of Committing to Market Segmentation

The key implication is that the organisation needs to commit to the segmentation strategy on the long term. There are costs of performing the research, fielding surveys, and focus groups, designing multiple packages, and designing multiple advertisements and communication messages. Potentially required changes include the development of new products, the modification of existing products, changes in pricing and distribution channels used to sell the product, as well as all communications with the market. These changes, in turn, are likely to influence the internal structure of the organisation, which may need to be adjusted in view of, for example, targeting a handful of different market segments.

Implementation Barriers:-

The first group of barriers relates to senior management. Lack of leadership, pro-active championing, commitment and involvement in the market segmentation process by senior leadership undermines the success of market segmentation.

A second group of barriers relates to organisational culture. Lack of market or consumer orientation, resistance to change and new ideas, lack of creative thinking, bad communication and lack of sharing of information and insights across organisational units, short-term thinking, unwillingness to make changes and office politics have been identified as preventing the successful implementation of market segmentation.

Most of these barriers can be identified from the outset of a market segmentation study, and then proactively removed. If barriers cannot be removed, the option of abandoning the attempt of exploring market segmentation as a potential future strategy should be seriously considered.

Step 2: Specifying the Ideal Target Segment

Segment Evaluation Criteria

In Step 2 the organisation must determine two sets of segment evaluation criteria. One set of evaluation criteria can be referred to as knock-out criteria. These criteria are the essential, non-negotiable features of segments that the organisation would consider targeting. The second set of evaluation criteria can be referred to as attractiveness criteria. These criteria are used to evaluate the relative attractiveness of the remaining market segments – those in compliance with the knock-out criteria.

The shorter set of knock-out criteria is essential. It is not up to the segmentation team to negotiate the extent to which they matter in target segment selection. The second, much longer and much more diverse set of attractiveness criteria represents a shopping list for the segmentation team. Members of the segmentation team need to select which of these criteria they want to use to determine how attractive potential target segments are. The segmentation team also needs to assess the relative importance of each attractiveness criterion to the organisation.

Knock-Out Criteria:-

Knock-out criteria are used to determine if market segments resulting from the market segmentation analysis qualify to be assessed using segment attractiveness criteria.

- The segment must be homogeneous; members of the segment must be similar to one another.
- The segment must be distinct; members of the segment must be distinctly different from members of other segments. 34 4 Step 2: Specifying the Ideal Target Segment
- The segment must be large enough; the segment must contain enough consumers to make it worthwhile to spend extra money on customising the marketing mix for them.
- The segment must be matching the strengths of the organisation; the organisation must have the capability to satisfy segment members' needs.
- Members of the segment must be identifiable; it must be possible to spot them in the marketplace.
- The segment must be reachable; there has to be a way to get in touch with members of the segment in order to make the customised marketing mix accessible to them.

Knock-out criteria must be understood by senior management, the segmentation team, and the advisory committee.

Attractiveness Criteria

Attractiveness criteria are not binary in nature. Segments are not assessed as either complying or not complying with attractiveness criteria. Rather, each market segment is rated; it can be more or less attractive with respect to a specific criterion. The attractiveness across all criteria determines whether a market segment is selected as a target segment.

Step 3: Collecting Data

Segmentation variable

The term segmentation variable to refer to the variable in the empirical data used in common-sense segmentation to split the sample into market segments. I

In commonsense segmentation, the segmentation variable is typically one single characteristic of the consumers in the sample. All the other personal characteristics available in the data – in this case: age, the number of vacations taken, and information about five benefits people seek or do not seek when they go on vacation – serve as so-called descriptor variables. They are used to describe the segments in detail. Describing segments is critical to being able to develop an effective marketing mix targeting the segment. The difference between common-sense and data-driven market segmentation is that data-driven market segmentation is based not on one, but on multiple segmentation variables. These segmentation variables serve as the starting point for identifying naturally existing, or artificially creating market segments useful to the organisation.

The same holds for data-driven market segmentation where data quality determines the quality of the extracted data-driven market segments, and the quality of the descriptions of the resulting segments. Good market segmentation analysis requires good empirical data.

Segmentation Criteria

The term segmentation variable refers to one measured value, for example, one item in a survey, or one observed expenditure category. The term segmentation criterion relates to the nature of the information used for market segmentation. It can also relate to one specific construct, such as benefits sought.

- **Geographic Segmentation :-** Geographic information is seen as the original segmentation criterion used for the purpose of market segmentation. The key advantage of geographic segmentation is that each consumer can easily be assigned to a geographic unit. As a consequence, it is easy to target communication messages, and select communication channels (such as local newspapers, local radio and TV stations) to reach the selected geographic segments.
- **Socio-Demographic Segmentation:-** Typical socio-demographic segmentation criteria include age, gender, income and education. Socio-demographic segments can be very useful in some industries. For example: luxury goods (associated with high income), cosmetics (associated with gender; even in times where men are targeted, the female and male segments are treated distinctly differently), baby

products (associated with gender), retirement villages (associated with age), tourism resort products (associated with having small children or not). As is the case with geographic segmentation, socio-demographic segmentation criteria have the advantage that segment membership can easily be determined for every consumer. In some instances, the socio-demographic criterion may also offer an explanation for specific product preferences (having children, for example, is the actual reason that families choose a family vacation village where previously, as a couple, their vacation choice may have been entirely different).

- **Psychographic Segmentation:** - When people are grouped according to psychological criteria, such as their beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product, the term psychographic segmentation is used. Psychographic criteria are, by nature, more complex than geographic or sociodemographic criteria because it is difficult to find a single characteristic of a person that will provide insight into the psychographic dimension of interest. As a consequence, most psychographic segmentation studies use a number of segmentation variables, for example: a number of different travel motives, a number of perceived risks when going on vacation.
- **Behavioural Segmentation:** - A wide range of possible behaviours can be used for this purpose, including prior experience with the product, frequency of purchase, amount spent on purchasing the product on each occasion (or across multiple purchase occasions), and information search behaviour. The key advantage of behavioural approaches is that – if based on actual behaviour rather than stated behaviour or stated intended behaviour – the very behaviour of interest is used as the basis of segment extraction. But behavioural data is not always readily available, especially if the aim is to include in the segmentation analysis potential customers who have not previously purchased the product, rather than limiting oneself to the study of existing customers of the organisation.

Data from Survey Studies

Most market segmentation analyses are based on survey data. Survey data is cheap and easy to collect, making it a feasible approach for any organisation. But survey data – as opposed to data obtained from observing actual behaviour – can be contaminated by a wide range of biases. A few key aspects that need to be considered when using survey data are discussed below.

- ### Choice of Variables

  In data-driven segmentation, all variables relevant to the construct captured by the segmentation criterion need to be included.

- **Response Options**

Options allowing respondents to answer in only one of two ways, generate binary or dichotomous data. Such responses can be represented in a data set by 0s and 1s. The distance between 0 and 1 is clearly defined and, as such, poses no difficulties for subsequent segmentation analysis. Preferably, therefore, either metric or binary response options should be provided to respondents if those options are meaningful with respect to the question asked. Using binary or metric response options prevents subsequent complications relating to the distance measure in the process of data-driven segmentation analysis.

- **Response Styles**

  A response bias is a systematic tendency to respond to a range of questionnaire items on some basis other than the specific item content. If a bias is displayed by a respondent consistently over time, and independently of the survey questions asked, it represents a response style.

- **Sample Size**

  Many statistical analyses are accompanied by sample size recommendations. Not so market segmentation analysis. When the sample size is insufficient, it is impossible to determine which the correct number of market segments is. If the sample size is sufficient, however it is very easy to determine the number and nature of segments in the data set. Viennese psychologist Formann (1984) recommends that the sample size should be at least 2p (better five times 2p), where p is the number of segmentation variables.

  Market characteristics studied included: the number of market segments present in the data, whether those market segments are equal or unequal in size, and the extent to which market segments overlap. Larger sample sizes always improve an algorithm's ability to identify the correct market segmentation solution. The extent to which this is the case, however, varies substantially across market and data characteristics. Also, some of the challenging market and data characteristics can be compensated by increasing sample size; others cannot. For example, using uncorrelated segmentation variables leads to very good segment recovery. But, correlation cannot be well compensated for by increasing sample size. Overall, this study demonstrates the importance of having a sample size sufficiently large to enable an algorithm to extract the correct segments (if segments naturally exist in the data).

It can be concluded from the body of work studying the effects of survey data quality on the quality of market segmentation results based on such data that, optimally, data used in market segmentation analyses should

- contain all necessary items;
- contain no unnecessary items;
- contain no correlated items;
- contain high-quality responses;
- be binary or metric;
- be free of response styles;
- include responses from a suitable sample given the aim of the segmentation study; and
- include a sufficient sample size given the number of segmentation variables (100 times the number of segmentation variables)

Sociodemographics		Travel behaviour	Benefits sought				
gender	age	Nº of vacations	relaxation	action	culture	explore	meet people
Female	34	2	1	0	1	0	1
Female	55	3	1	0	1	0	1
Male	87	2	1	0	1	0	1
Female	68	1	0	1	1	0	0
Female	34	1	0	0	1	0	0
Female	22	0	1	0	1	1	1
Female	31	3	1	0	1	1	1
Male	55	4	0	1	0	1	1
Male	43	0	0	1	0	1	0
Male	23	0	0	1	1	0	1
Male	19	3	0	1	1	0	1
Male	64	4	0	0	0	0	0
descriptor variables			segmentation variables				

Data from Internal Sources

When organisations rely solely on internal data, they risk limiting their understanding of the market to their existing customers, while ignoring potential customers who may have different needs and preferences. This can result in a biased market segmentation analysis that does not accurately represent the broader market. Therefore, it's important for organisations to supplement internal data with external data sources, such as market research surveys, to gain a more comprehensive understanding of the market and its potential customers. Additionally, it's important for organisations to carefully consider the potential biases in their internal data and take steps to mitigate them, such as by weighting the data to account for the over-representation of existing customers.

Data from Experimental Studies

Experimental data can provide a valuable source of information for market segmentation analysis. Such data can be collected from field or laboratory experiments that test how people respond to certain advertisements. The response to these advertisements can then be used as a segmentation criterion. Additionally, experimental data can result from choice experiments or conjoint analyses, which present consumers with carefully developed stimuli consisting of specific levels of product attributes. Consumers indicate which products they prefer, providing information about the extent to which each attribute and attribute level affects choice, which can also be used as a segmentation criterion.

STEP Selecting the Target Segment(s)

The Targeting Decision

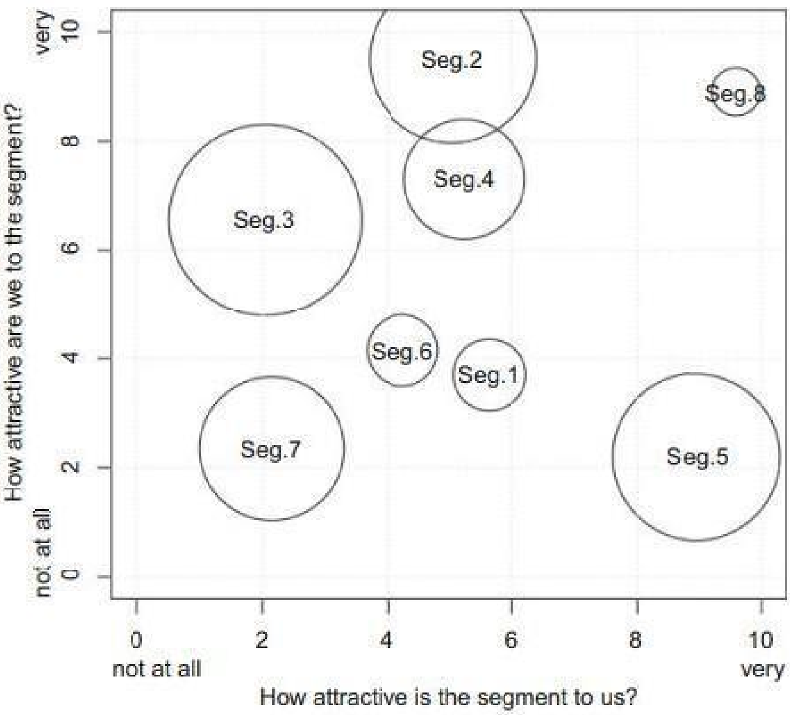
Market segmentation involves the selection of one or more target segments for an organization. This is a long-term decision that significantly affects the organization's future performance. Before selecting a target segment, knock-out criteria for market segments have been agreed upon, and segment attractiveness criteria have been selected and weighed. It is to ensure that all the market segments still under consideration have passed the knock-out criteria test. Then, the attractiveness of the remaining segments and the relative organizational competitiveness for these segments needs to be evaluated. The target segment decision is based on answering two broad categories of questions: which segment the organization most wants to target, and which organization offering the same product each segment would most like to buy from and commit to.

Market Segment Evaluation

market segments based on their attractiveness and relative organizational competitiveness. This can be done using a generic segment evaluation plot that measures the two criteria along the axes. The article also highlights the importance of returning to the specifications of an ideal target segment and assigning a value for each attractiveness criterion to each segment. This information is critical for target segment selection in market segmentation analysis.

The article discusses how to evaluate market segments based on their attractiveness and relative organizational competitiveness. This can be done using a generic segment evaluation plot that measures the two criteria along the axes. The article also highlights the importance of returning to the specifications of an ideal target segment and assigning a value for each attractiveness criterion to each segment. This information is critical for target segment selection in market segmentation analysis.

Anything can be plotted onto the bubble size. Typically profit potential is plotted. Profit combines information about the size of the segment with spending and, as such, represents a critical value when target segments are selected. In other contexts, entirely different criteria may matter. For example, if a non for profit organisation uses market segmentation to recruit volunteers to help with land regeneration activities, they may choose to plot the number of hours volunteered as the bubble size.


Seg.1
Seg.2
Seg.3
Seg.4
Seg.5
Seg.6
Seg.7
Seg.8
0
2
4
6
8
10
not at all
very
How attractive is the segment to us?
How attractive are we to the segment?

<u>STEP 9:- Customising the Marketing Mix</u>

Implications for Marketing Mix Decisions

Market segmentation does not stand independently as a marketing strategy. Rather, it goes hand in hand with the other areas of strategic marketing, most importantly: positioning and competition. In fact, the segmentation process is frequently seen as part of what is referred to as the segmentation-targeting-positioning (STP) approach (Lilien and Rangaswamy 2003). The segmentation-targeting-positioning approach postulates a sequential process. The process starts with market segmentation (the extraction, profiling and description of segments), followed by targeting (the assessment of segments and selection of a target segment), and finally positioning (the measures an organisation can take to ensure that their product is perceived as distinctly different from competing products, and in line with segment needs).

Viewing market segmentation as the first step in the segmentation-targetingpositioning approach is useful because it ensures that segmentation is not seen as independent from other strategic decisions. It is important, however, not to adhere too strictly to the sequential nature of the segmentation-targeting-positioning process. It may well be necessary to move back and forward from the segmentation to the targeting step, before being in the position of making a long-term commitment to one or a small number of target segments.

How the target segment decision affects marketing mix development

How the target segment decision – which has to be integrated with other strategic areas such as competition and positioning – affects the development of the marketing mix. For reasons of simplicity, the traditional 4Ps model of the marketing mix including Product, Price, Place and Promotion serves as the basis of this discussion. Be it twelve or four, each one of those aspects needs to be thoroughly reviewed once the target segment or the target segments have been selected. To best ensure maximising on the benefits of a market segmentation strategy, it is important to customise the marketing mix to the target segment .

The selection of one or more specific target segments may require the design of new, or the modification or re-branding of existing products (Product), changes to prices or discount structures (Price), the selection of suitable distribution channels (Place), and the development of new communication messages and promotion strategies that are attractive to the target segment (Promotion). One option available to the organisation is to structure the entire market segmentation analysis around one of the 4Ps. This affects the choice of segmentation variables. If, for example, the segmentation analysis is undertaken to inform pricing decisions, price sensitivity, deal proneness, and price sensitivity represent suitable segmentation variables.

If the market segmentation analysis is conducted for the purpose of informing distribution decisions, store loyalty, store patronage, and benefits sought when selecting a store may represent valuable segmentation variables. Typically, however, market segmentation analysis is not conducted in view of one of the 4Ps specifically. Rather, insights gained from the detailed description of the target segment resulting from guide the organisation in how to develop or adjust the marketing mix to best cater for the target segment chosen.

Product

One of the key decisions an organisation needs to make when developing the product dimension of the marketing mix, is to specify the product in view of customer needs. Often this does not imply designing an entirely new product, but rather modifying an existing one. Other marketing mix decisions that fall under the product dimension are: naming the product, packaging it, offering or not offering warranties, and after sales support services.

In terms of the product targeted at this market segment, possible product measures may include developing a new product. For example, a MUSEUMS, MONUMENTS & MUCH, MUCH MORE product (accompanied by an activities pass) that helps members of this segment to locate activities they are interested in, and points to the existence of these offers at the destination during the vacation planning process. Another opportunity for targeting this segment is that of proactively making gardens at the destination an attraction in their own right.

Price

Typical decisions an organisation needs to make when developing the price dimension of the marketing mix include setting the price for a product, and deciding on discounts to be offered.

Place

The key decision relating to the place dimension of the marketing mix is how to distribute the product to the customers. This includes answering questions such as: should the product be made available for purchase online or offline only or both; should the manufacturer sell directly to customers; or should a wholesaler or a retailer or both be used. Returning to the example of members of segment 3 and the destination with a rich cultural heritage: the survey upon which the market segmentation analysis was based also asked survey respondents to indicate how they booked their accommodation during their last domestic holiday. Respondents could choose multiple options. This information is place valuable; knowing the booking preferences of members of segment 3 enables the destination to ensure that the MUSEUMS, MONUMENTS & MUCH, MUCH MORE product is bookable through these very distribution channels.

Promotion

Typical promotion decisions that need to be made when designing a marketing mix include: developing an advertising message that will resonate with the target market, and identifying the most effective way of communicating this message. Other tools in the promotion category of the marketing mix include public relations, personal selling, and sponsorship.