

Chapter 5

Step 3: Collecting Data



5.1 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. 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. 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.

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 behaviour, 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 organisation. An illustration is provided in Table 5.2 using the same data as in Table 5.1.

Table 5.1 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	
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	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					

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) characterised 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 customised 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 behaviour. Survey data – although it arguably represents the most common source of data for market segmentation studies – can be unreliable in reflecting behaviour, especially when the behaviour 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 behaviour is preferable.

5.2 Segmentation Criteria

Long before segments are extracted, and long before data for segment extraction is collected, the organisation 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 behavioural.

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.

5.2.1 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. While simple, the geographic segmentation approach is often the most appropriate. For example: if the national tourism organisation of Austria wants to attract tourists from neighbouring 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 neighbouring 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 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.

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. While, for example, people residing in luxury suburbs may all be a good target market for luxury cars, location is rarely the reason for differences in product preference. 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.

5.2.2 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 behaviour. 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.

5.2.3 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 behaviour. 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.

5.2.4 Behavioural Segmentation

Another approach to segment extraction is to search directly for similarities in behaviour or reported behaviour. 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. In a comparison of different segmentation criteria used as segmentation variables, behaviours reported by tourists emerged as superior to geographic variables (Moscardo et al. 2001).

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. As such, behavioural 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 behaviour 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 behavioural data also avoids the need for the development of valid measures for psychological constructs.

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.

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

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 (Nunnally 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 categorised 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.

5.3.2 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, respondents 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. Although ordinal scales dominate both market research and academic survey research, using binary or metric response options instead is usually not a compromise. If, for example, there is a strong reason to believe that very fine nuances of responses need to be captured, and if capturing those fine nuances does not come at the cost of also capturing response styles, this can be achieved using visual analogue scales. 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), especially when formulated in a level free way (see the discussion of the doubly level free answer format with individually inferred thresholds, or DLF IIST, in Rossiter et al. 2010; Rossiter 2011; Dolnicar and Grün 2013).

5.3.3 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.

5.3.4 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.

Only a small number of studies have investigated this problem. Viennese psychologist Formann (1984) recommends that the sample size should be at least 2^p (better five times 2^p), where p is the number of segmentation variables. This rule of thumb relates to the specific purpose of goodness-of-fit testing in the context of latent class analysis when using binary variables. It can therefore not be assumed

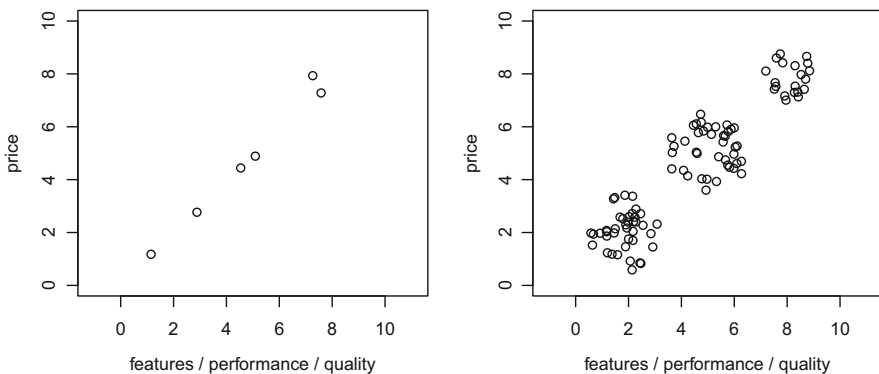


Fig. 5.1 Illustrating the importance of sufficient sample size in market segmentation analysis

to be generalisable to other algorithms, inference methods, and scales. Qiu and Joe (2015) developed a sample size recommendation for constructing artificial data sets for studying the performance of clustering algorithms. According to Qiu and Joe (2015), the sample size should – in the simple case of equal cluster sizes – be at least ten times the number of segmentation variables times the number of segments in the data ($10 \cdot p \cdot k$ where p represents the number of segmentation variables and k represents the number of segments). If segments are unequally sized, the smallest segment should contain a sample of at least $10 \cdot p$.

Dolnicar et al. (2014) conducted extensive simulation studies with artificial data modelled after typical data sets used in applied tourism segmentation studies. Knowing the true structure of the data sets, they tested sample size requirement for algorithms to correctly identify the true segments. Figure 5.2 shows the effect of sample size on the correctness of segment recovery for this particular study. The adjusted Rand index serves as the measure of correctness of segment recovery. The adjusted Rand index assesses the congruence between two segmentation solutions. Higher values indicate better alignment. Its maximum possible value is 1. The expected value is 0 if the two segmentation solutions are derived independently in a random way. To assess segment recovery, the adjusted Rand index is calculated for the true segment solution and the extracted one.

In Fig. 5.2, 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.

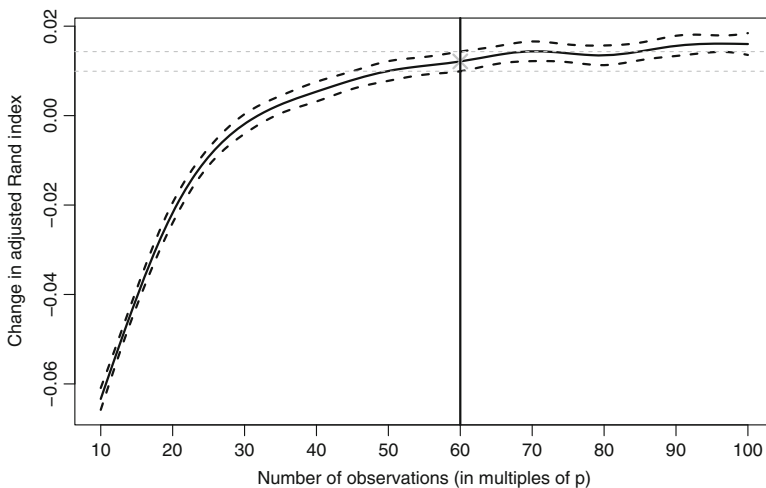


Fig. 5.2 Effect of sample size on the correctness of segment recovery in artificial data. (Modified from Dolnicar et al. 2014)

Not surprisingly, increasing the sample size improves the correctness of the extracted segments. Interestingly, however, the biggest improvement is achieved by increasing very small samples. As the sample size increases, the marginal benefit of further increasing the sample size decreases. Based on the results shown in Fig. 5.2, a sample size of at least $60 \cdot p$ is recommended. For a more difficult artificial data scenario Dolnicar et al. (2014) recommend using a sample size of at least $70 \cdot p$; no substantial improvements in identifying the correct segments were identified beyond this point.

Dolnicar et al. (2016) extended this line of research to account for key features of typical survey data sets, making it more difficult for segmentation algorithms to identify correct segmentation solutions. Specifically, they investigated the effect on sample size requirements resulting from market characteristics not under the control of the data analyst and, data characteristics – at least to some degree – under the control of the data analyst.

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. De Craen et al. (2006) show that the presence of unequally sized segments makes it more difficult for an algorithm to extract the correct market segments. Steinley (2003) shows the same for the case of overlapping segments.

In addition, some of the characteristics of survey data discussed above have been shown to affect segment recovery, specifically: sampling error, response biases and response styles, low data quality, different response options, the inclusion of irrelevant items, and correlation between blocks of items. Figure 5.3 shows the

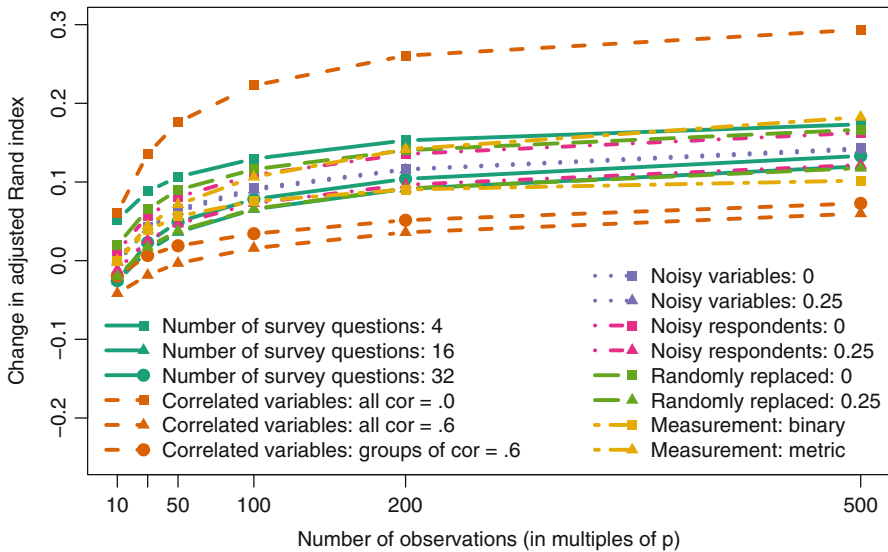


Fig. 5.3 Sample size requirements in dependence of market and data characteristics. (Modified from Dolnicar et al. 2016)

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.

As can be seen in Fig. 5.3, 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, as can be seen in Fig. 5.3: the top-most and the two bottom-most curves in Fig. 5.3 show three different levels of correlation between segmentation variables. If the variables are not correlated at all, the algorithm has no difficulty extracting the correct segments. If, however, the variables are highly correlated, the task becomes so difficult for the algorithm, that even increasing the sample size dramatically does not help. A small number of noisy variables, on the other hand, has a lower effect.

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). The recommendation by Dolnicar et al. (2016) is to ensure the data contains at least 100 respondents for each segmentation variable. Results from this study also highlight the importance of collecting high-quality unbiased data as the basis for market segmentation analysis.

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).

5.4 Data from Internal Sources

Increasingly organisations have access to substantial amounts of internal data that can be harvested for the purpose of market segmentation analysis. Typical examples are scanner data available to grocery stores, booking data available through airline loyalty programs, and online purchase data. The strength of such data lies in the

fact that they represent *actual* behaviour of consumers, rather than statements of consumers about their behaviour or intentions, known to be affected by imperfect memory (Niemi 1993), as well as a range of response biases, such as social desirability bias (Fisher 1993; Paulhus 2002; Karlsson and Dolnicar 2016) or other response styles (Paulhus 1991; Dolnicar and Grün 2007a,b, 2009).

Another advantage is that such data are usually automatically generated and – if organisations 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. What is missing is information about other consumers the organisation may want to win as customers in future, which may differ systematically from current customers in their consumption patterns.

5.5 Data from Experimental Studies

Another possible source of data that can form the basis of market segmentation analysis is experimental data. 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. The aim of such studies is to present consumers with carefully developed stimuli consisting of specific levels of specific product attributes. Consumers then indicate which of the products – characterised by different combinations of attribute levels – they prefer. Conjoint studies and choice experiments result in information about the extent to which each attribute and attribute level affects choice. This information can also be used as a segmentation criterion.

5.6 Step 3 Checklist

Task	Who is responsible?	Completed?
Convene a market segmentation team meeting.		<input type="checkbox"/>
Discuss which consumer characteristics could serve as promising segmentation variables. These variables will be used to extract groups of consumers from the data.		<input type="checkbox"/>
Discuss which other consumer characteristics are required to develop a good understanding of market segments. These variables will later be used to describe the segments in detail.		<input type="checkbox"/>
Determine how you can collect data to most validly capture both the segmentation variables and the descriptor variables.		<input type="checkbox"/>
Design data collection carefully to keep data contamination through biases and other sources of systematic error to a minimum.		<input type="checkbox"/>
Collect data.		<input type="checkbox"/>

References

Bieger T, Laesser C (2002) Market segmentation by motivation: the case of Switzerland. *J Travel Res* 41(1):68–76

Bock T, Uncles M (2002) A taxonomy of differences between consumers for market segmentation. *Int J Res Mark* 19(3):215–224

Bockenholt U, Langeheine UR (1996) Latent change in recurrent choice data. *Psychometrika* 61(2):285–301

Boksberger PE, Laesser C (2009) Segmentation of the senior travel market by the means of travel motivations. *J Vacat Mark* 15(4):311–322

Brusco M (2004) Clustering binary data in the presence of masking variables. *Psychol Methods* 9(4):510–523

Cahill DJ (2006) *Lifestyle market segmentation*. Haworth Press, New York

Carmone FJ, Kara A, Maxwell S (1999) HINoV: a new model to improve market segment definition by identifying noisy variables. *J Mark Res* 36(4):501–509

Churchill JGA (1979) A paradigm for developing better measures of marketing constructs. *J Mark Res* 16:64–73

De Craen S, Commandeur JJF, Frank LE, Heiser WJ (2006) Effects of group size and lack of sphericity on the recovery of clusters in k-means cluster analysis. *Multivar Behav Res* 41(2):127–145

DeSarbo W, Mahajan V (1984) Constrained classification: the use of a priori information in cluster analysis. *Psychometrika* 49(2):187–215

- DeSarbo WS, Carroll JD, Clark LA, Green PE (1984) Synthesized clustering: a method for amalgamating alternative clustering bases with differential weighting of variables. *Psychometrika* 49(1):57–78
- Dolnicar S (2003) Simplifying three-way questionnaires – do the advantages of binary answer categories compensate for the loss of information? In: ANZMAC CD proceedings, Australia and New Zealand marketing academy (ANZMAC)
- Dolnicar S, Grün B (2007a) Assessing analytical robustness in cross-cultural comparisons. *Int J Tour Cult Hosp Res* 1(2):140–160
- Dolnicar S, Grün B (2007b) Cross-cultural differences in survey response patterns. *Int Mark Rev* 24(2):127–143
- Dolnicar S, Grün B (2009) Response style contamination of student evaluation data. *J Mark Educ* 31(2):160–172
- Dolnicar S, Grün B (2013) Validly measuring destination images in survey studies. *J Travel Res* 52(1):3–13
- Dolnicar S, Rossiter JR (2008) The low stability of brand-attribute associations is partly due to measurement factors. *Int J Res Mark* 25(2):104–108
- Dolnicar S, Grün B, Leisch F (2011) Quick, simple and reliable: forced binary survey questions. *Int J Mark Res* 53(2):231–252
- Dolnicar S, Rossiter JR, Grün B (2012) “Pick-any” measures contaminate brand image studies. *Int J Mark Res* 54(6):821–834
- Dolnicar S, Grün B, Leisch F, Schmidt K (2014) Required sample sizes for data-driven market segmentation analyses in tourism. *J Travel Res* 53(3):296–306
- Dolnicar S, Grün B, Leisch F (2016) Increasing sample size compensates for data problems in segmentation studies. *J Bus Res* 69:992–999
- Fisher RJ (1993) Social desirability bias and the validity of indirect questioning. *J Consum Res* 20:303–315
- Formann A (1984) *Die Latent-Class-Analyse: Einführung in die Theorie und Anwendung*. Beltz, Weinheim
- Haley R (1968) Benefit segmentation: a decision-oriented tool. *J Mark* 32:30–35
- Haley RI (1985) *Developing effective communications strategy – a benefit segmentation approach*. Wiley, New York
- Haverila MJ (2013) Market segmentation in the cell phone market among adolescents and young adults. *Asia Pac J Mark Logist* 25(3):346–368
- Heilman CM, Bowman D (2002) Segmenting consumers using multiple-category purchase data. *Int J Res Mark* 19(3):225–252
- Hoek J, Gendall P, Esslemont D (1996) Market segmentation: a search for the holy grail? *J Mark Pract Appl Mark Sci* 2(1):25–34
- Johnson M, Lehmann D, Horne D (1990) The effects of fatigue on judgments of interproduct similarity. *Int J Res Mark* 7(1):35–43
- Karlsson L, Dolnicar S (2016) Does eco certification sell tourism services? Evidence from a quasi-experimental observation study in Iceland. *J Sustain Tour* 24(5):604–714
- Laesser C, Crouch GI, Beritelli P (2006) Market segmentation by reasons and influences to visit a destination: the case of international visitors to Australia. *Tour Anal* 11(4):241–249
- Lewis RC, Chambers RE, Chacko HE (1995) *Marketing leadership in hospitality: foundations and practices*. van Nostrand Reinhold, New York
- Milligan GW (1980) An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* 45(3):325–342
- Moscardo G, Pearce P, Morrison A (2001) Evaluating different bases for market segmentation. *J Travel Tour Mark* 10(1):29–49
- Niemi I (1993) Systematic error in behavioural measurement: comparing results from interview and time budget studies. *Soc Indic Res* 30(2–3):229–244
- Nunnally J (1978) *Psychometric theory*, 3rd edn. McGraw Hill, New York

- Paulhus D (1991) Measurement and control of response bias. In: Robinson J, Shaver P, Wrightsman L (eds) Measures of personality and social psychological attitudes. Academic Press, San Diego, pp 17–59
- Paulhus DL (2002) Socially desirable responding: the evolution of a construct. In: Braun HI, Jackson DN, Wiley DE (eds) The role of constructs in psychological and educational measurement, chap 4. Routledge, Lawrence Erlbaum Associates Mahwah, New Jersey, London pp 51–76
- Poulsen CS (1990) Mixed Markov and latent Markov modelling applied to brand choice behaviour. *Int J Res Mark* 7:5–19
- Qiu W, Joe H (2015) clusterGeneration: random cluster generation (with specified degree of separation). <https://CRAN.R-project.org/package=clusterGeneration>, R package version 1.3.4
- Ramaswamy V (1997) Evolutionary preference segmentation with panel survey data: an application to new products. *Int J Res Mark* 14(1):57–89
- Rossiter JR (2002) The C-OAR-SE procedure for scale development in marketing. *Int J Res Mark* 19(4):305–335
- Rossiter JR (2011) Measurement for the social sciences. The C-OAR-SE method and why it must replace psychometrics. Springer, New York
- Rossiter JR, Dolnicar S, Grün B (2010) The LFFB comparative judgment measure of brand attribute beliefs. Working Paper, Faculty of Commerce, University of Wollongong
- Steenkamp JBEM, Ter Hofstede F (2002) International market segmentation: issues and perspectives. *Int J Res Mark* 19(3):185–213
- Steinley D (2003) Local optima in k-means clustering: what you don't know may hurt you. *Psychol Methods* 8(3):294–304
- Tsai CY, Chiu CC (2004) A purchase-based market segmentation methodology. *Exp Syst Appl* 27(2):265–276
- Tynan AC, Drayton J (1987) Market segmentation. *J Mark Manag* 2(3):301–335
- Yankelovich D, Meer D (2006) Rediscovering market segmentation. *Harv Bus Rev* 84(2):122–131

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

