Matkailututkimus 11: 2, 8-24 (2014)
©Suomen matkailututkimuksen seura

Using cluster analysis to segment tourists: response-style effects

Juho Pesonen and Antti Honkanen
University of Eastern Finland

Different people have different response styles, some favour high scores whereas others low scorers and some people avoid using extreme scores, especially in attitudinal research questions. This is called response style effect. The goal of segmentation is to group people into homogenous clusters that differ from each other significantly. According to a literature review of earlier segmentation studies, there is a possibility that response style effects can influence segmentation results. In this study, empirical data is used to examine how response style effects affect cluster formation in a single country context using k-means cluster analysis. Results show that especially when using cluster analysis, response style effects should be closely examined as the results can be drastically different.

Keywords: segmentation, rural tourism, methodology, response style effects, cluster analysis, tourism

Introduction

Segmentation has been used for decades to find customer groups best suited for different companies Segmentation is also a very popular topic in academic literature. For example Dolnicar (2006a) reviewed 75 tourism segmentation manuscripts, of which most were published in the Journal of Travel Research, Tourism Management and the Journal of Travel and Tourism Marketing from 1981 until 2005. This review demonstrates the wide acceptance and use of market segmentation in tourism literature. However, market segmentation has several pitfalls for researches especially regarding to methodology. According to Dolnicar (2006a), most popular method of segmentation in the reviewed studies was K-means cluster analysis. This study shows why it is critical to carefully examine response style effects when conducting market segmentation studies with K-means cluster analysis, something that is often neglected in market segmentation in tourism.

According to Hair et al. (2010), there is a possibility that when clustering data collected using for example a number of ratings on a 10-point scale, we could get clusters of respondents for whom everything was important, clusters where everything had little importance and maybe some clusters in between. This is called response-style effect and results resembling it can be found in many different segmentation studies in tourism. There are many reasons why response style effects occur, including culture and socio-demographics (De Jong et al., 2008). Typically culture-specific response style effects are ignored in research even in multi-national studies (Dolnicar et al., 2008). When all the respondents are from the same country response style effects are hardly ever examined.

The goal of this study is to show how different and commonly used data analysis methods in tourism segmentation results in different segments and lead to different conclusions when response style effects are taken into account. This study examines and demonstrates how response style effects affect cluster analysis results even in a single nationality context. This research contributes to market segmentation practices by providing evidence on the importance of response style effects for cluster analysis. Also advice on how researchers should take response style effects into account when cluster analysis is used to segment tourists are given.

According to Dolnicar and Grün (2007), the existence of response styles is a well-known and much-studied phenomenon in various disciplines within the empirical social sciences. Respondents have a tendency to answer in different ways to answer formats that researchers offer them. Some respondents make much more use of extreme answer options whereas other respondents feel more comfortable avoiding extreme answers and make more use of the middle answer categories for example in answer formats based on a Likert scale. Very often in tourism literature, response styles are linked to respondents from different cultures. Understanding the effects of nationality can also be regarded as very important in studying response styles. In tourism studies, multinational data is often used and examining response styles between respondents from different countries should be automatic. However, there is not much information on what role response styles play when respondents are from one country.

Success of segmentation is very much dependent of the segmentation solution. According to Middleton and Clarke (2001), segments are assumed to have homogenous travel behaviours and there should also be meaningful differences among the segments within a total market. Sally Dibb (1998) states that segmentation failure occurs when it has not been possible to use the segmentation to develop a suitable marketing mix. This means that segments should be distinctive from each other. If cluster analysis groups respondents with similar answering styles into same segments, the segments are based on response styles instead of actual market structures. This means that the results are not very useful for marketing management purposes.

This study aims to examine response style effects when using k-means cluster analysis to find segmentation solution and find out how different response styles affect cluster analysis results in a single country context. The study is structured into four parts after the introduction. In the first part, a literature review of segmentation and response style effects in tourism studies is conducted. In the second part, the methods and the data used in this study are explained in detail. In the third part, results of the study are presented and in the fourth part, the results are discussed and limitations and directions for further research presented.

Literature review

Response style effects

Response style effects have been well-known for over 30 years. Response style refers to individual differences in respondents to distribute responses on rating scales. Baumgartner and Steencamp (2001) define response style effects as tendencies to respond systematically to questionnaire items on some basis other than what the items were specifically designed to measure. The second definition refers to response bias. Sometimes terms answer format effects, response set bias and anchoring point difference are used in the same meaning as response style effects. In this article, the term response style effect is used to describe the phenomenon.

There are several different response styles. The most common types are extreme response style (ERS) and acquiescence response style (ARS). Using mostly the extremes on Likert scales indicates ERS. Using the positive range of the scale and avoiding negative ones can be a sign of ARS. Some kind of reversed ERS is at issue when respondents are avoiding the endpoints and responses are concentrated toward the middle of the scale. The style is called midpoint responding (MPR). Other types of response style effects are disacquiescence response style (DARS), net acquiescence response style (NARS), response range (RR) and non-contingent responding (NCR). In this article only ARS and ERS are taken under closer examination as a simple approach for examining ARS and ERS is presented by Dolnicar and Grün (2009).

Response style effects involve many problems. ERS, for example, spuriously increase reliability but decrease validity. It also produces a frequency distribution with more extreme values. If correlation based analysis is used, increased standard deviations and decreased correlations distort results (Dolnicar & Grün, 2009).

In factor analysis, ARS may cause an additional factor with negative loadings. The presence of it is a reason to check ARS but there are other possible reasons for a factor with negative loadings. Respondents preferring either upper or lower answer options lead to higher but false correlations. In general, it can be said that response style effects lead to biased results in many analyses based on correlations or similarities and dissimilarities in a data matrix. Cluster analysis is one example of this kind of data analysis.

Dolnicar and Grün (2009) propose a 3-step procedure for identifying potential response style bias. In step one, the existence of empirical evidence of response style effects is checked. The use of prior knowledge is recommended. Sample consisting of respondents from for example different cultures indicates the possibility of response style effects.

Response style effects in tourism have been traditionally examined in cross-cultural context. Dolnicar and Grün (2007) state respondents from different cultural backgrounds tend to use survey answer formats in different ways, but the effect does not influence the results of empirical studies within one discrete cultural area. For instance, African-American and Hispanic respondents tend to give more extreme answers whereas Asians use the extreme option less (Dolnicar 2006b, pp. 201–202). However, this brings forth a question of what is a discrete cultural area? Traditionally in tourism segmentation studies countries have been regarded as single cultural areas (see Dolnicar and Grün, 2007), thus regarding respondents from the same country unaffected by response styles.

When thinking of for example the USA, it seems clear that the culture of respondents from rural Midwest differ greatly from that of someone living in the heart of New York City. Same is also true for Finland in many different ways, for example Western Finland has been influenced by the proximity of Sweden, whereas in Eastern Finland, Russia has had more influence. Therefore, this study aims to find if response style effects should be examined also in the context of a single nationality.

Culture is not the only source of response style but also other variables affect it. Weijters (2006), for example, found that demographics partially explained response styles but the data collection method also had a strong influence. However, the results of prior research have not been very consistent in studying the effects of socio-demographics on response styles (De Jong et al., 2008). When data collected by paper and pencil, telephone and online were compared the major finding was that the telephone data showed a lower level of MRS and a higher level of ARS (Weijters, 2006).

Response style effects can be also measured. Individual standard deviations (Chun et al., 1974) or the proportion of extreme responses in a Likert scale (Baumgartner et al., 2001) can be used a measure of ERS. However, Greenleaf (1992) suggests that the items should be uncorrelated and have equal extreme response proportions. In addition, the mean response to an item should be close to the midpoint of the scale. Response sets should be particularly designed for revealing response style effects, which is not possible in many cases for practical reasons such as limited space in the questionnaire form and research goals.

ARS can be measured by using proportion of (dis)agreements but also using other measures (Weijters, 2006). Baumgartner and Steenkamp (2001) recommend measuring ARS as the extent of agreement with items that are heterogeneous in content and the extent of agreement with both positively and negatively worded items within the same scale. The individual means are also used to detect ARS (Hui & Triandis, 1989). ARS measures set a special condition for questionnaire design. Because the prime goal is to define segments, it might not be possible to design a segmentation questionnaire to meet requirements for detecting response style effect.

If there is a reason to believe that response style effects exist, step two is to choose which methods are suitable for correcting it in the data (Dolnicar and Grün, 2009). The problem of data correction is that no guarantee exists if the procedure eliminates response style contamination. It is even possible that the modification leads to the further bias in the data. One possibility is to test different correction methods and compare the results. Typically, data is corrected for ARS using individual means as well as using country-specific means if the bias is supposed to be a result of respondents belonging to different nationalities. For ERS, the correction is made using individual standard deviations as well as using country-specific standard deviations. Often data is corrected for both ARS and ERS using the individual measures as well as the country-specific ones (Dolnicar et al., 2008).

When correction methods are used, step three is to compare the results of separate actual data analyses with original data and corrected data sets (Dolnicar and Grün, 2009). If analyses lead to different conclusions, response style effects have probably biased the results. There is not an exact method of choosing which one is a bias free result but some discussion on the topic can be found. Rosmalen et al. (2010) presented a latent-class bilinear multinomial logit model to identify response styles in cross-national segmentation context. They were able to graphically distinguish the effects on the response behaviour of the characteristics of a respondent and the content of an item. Rosmalen et al. (2010) found altogether 11 response styles in their data, demonstrating the importance of response style correlations in segmentation. If analysis with original data produces segmentation solution based on response style effects and standardized data produces a cluster solution with theoretically meaningful interpretable clusters, the latter should be used.

Response style effects in earlier tourism market segmentation studies

Segmentation has been used for decades to understand customer behaviour and help companies to promote and design their products. In tourism marketing, segmentation has been widely used and studied (Dolnicar, 2002). Most studies that have segmented customers based on data-driven methods have used a technique belonging to the family of cluster analysis (Everitt, 1993). According to Dolnicar (2002), cluster analysis is a toolbox of highly interdisciplinary techniques of multivariate data analysis and it has been used in various disciplines of social and natural sciences.

According to Hair et al. (2010), k-means are nonhierarchical clustering algorithms that work by portioning the data into a user-specified number of clusters and then iteratively reassign observations to clusters until some numerical criterion is met. According to Dolnicar (2002), k-means is the most popular algorithm among partitioning approaches when using cluster analysis in tourism. K-means cluster analysis is seldom used without data pre-processing. Most popular methods of data pre-processing are factor analysis and standardization. According to Dolnicar (2002), standardization tends to lead to a distortion of results, as actually existing clusters are hidden and instead clusters in a transformed space are searched for. The common use of factor analysis before clustering is also a questionable standard (Dolnicar, 2002).

In some tourism segmentation studies, the results are very similar to the description of bias caused by response-style effects. Several authors (e.g. Calantone et al., 1978; Alpert, 1980) have warned about response style effects when cluster analyses are used for identifying market segments.

Bieger and Laesser (2002) segmented Swiss tourists according to their motivations. Ten motivation items were measured using four-point Likert scale. Bieger and Laesser (2002) used k-means cluster analysis to form groups from the entire sample using individual responses on the basis of motivations. They found four different segments, of which one had no highest motivations among all clusters and one had seven out of ten. The cluster with no highest motivations had eight lowest motivations and cluster with seven highest motivations had no lowest motivations. This suggests that clustering is at least partly based on response style and especially extreme response style.

Chung et al. (2004) conducted a benefit segmentation of hotel guests in deluxe hotels in Seoul. They used a multinational sample and measured hotel attributes, i.e. benefits, with a five-point Likert scale. They used principal component analysis with varimax rotation to examine underlying dimensions and clustered hotel visitors with quick cluster analysis. They found five clusters, of which one was labelled as Seek-few benefits and one as Seek-all-benefits. "Seek-few benefits" had very many low scores whereas "Seek-all-benefits" had many high scores. Most of the respondents in "Seek-all-benefits" were American and 50 % of Asian respondents of their study belong to "Seek-few benefits" cluster. The presence of ARS can be suspected.

Füller and Matzler (2008) collected an online sample of ski resort customers. They measured 34 satisfaction items and ten lifestyle items using a five-point Likert scale. They obtained five cluster solutions using hierarchical clustering method and k-means cluster analysis. Cluster labelled as "Family" had lowest scores in nine lifestyle categories whereas cluster labelled as "Demanding" had highest scores in all lifestyle items. This pattern can also be seen from satisfaction scores: "Demanding" segment has agreement scores above sample average and "Family" segment has only scores below sample average.

May et al. (2001) segmented snowmobilers in Wyoming, USA, using Recreation Experience Preference (REP) scales. They were asked to rate the importance of 26 reasons for snowmobiling measured using seven-point scale ranging from extremely unimportant to extremely important. They used a number of different clustering techniques in conjunction with each other to classify the data of the study and found five clusters. Second cluster valued all the REP items more than other segments, whereas fourth valued everything except for one item less than other segments.

Last example of studies that could possibly be affected by response style effects was conducted by Park and Yoon (2009). They segmented rural tourists in Korea according to tourists' motivations. They used principal component analysis to identify the underlying motive dimensions. Then hierarchy cluster analysis and k-means cluster analysis were used to identify the number of clusters and to classify the samples according to their travel experience parameters that discriminated the tourists the best. They also state that they clustered individuals "in such a way that those within each cluster were more similar to each other than to those in other clusters, thereby creating a situation of homogeneity within clusters and heterogeneity between clusters" (Park and Yoon, 2009, p. 102). As a result, they found four segments of rural tourists, of which two were named as "Passive tourists" and "Want-it-all tourists". "Passive tourists" were tourists that valued everything lower than other tourists and "Want-it-all tourists" valued all the measured motivation factors more. Other two segments were in between these two. In the aforementioned studies the possibility of response style effects is not even discussed, even though it is important when using measures such as Likert scale and respondents come from several different cultures (Clarke, 2001; Dolnicar and Grün, 2007).

However, there are also many studies using similar methods that have been able to find distinctive segments. Beh and Bruyere (2007) used an ipsative clustering method based on the standardizing of the motivation factor scores to account for any potential individual response patterns in segmenting visitors in three Kenyan national reserves. All the three segments Beh and Bruyere (2007) found were different from each other, each regarding some motivations more important than other segments.

Frochot (2005) segmented tourists in rural areas in Scotland according to benefits sought. She measured 13 different benefit statements using five-point scale ranging from very important to unimportant. By analyzing the responses using factor analysis with Varimax rotation and k-means cluster analysis she was able to find four segments. None of these segments had had highest or lowest scores in all factors meaning that they differed from each other in what they value.

There are also a variety of other methodologies to do segmentation research in tourism. Beritelli et al. (2007) used combined two-step market segmentation of a priori and a posteriori segmentation by first dividing respondents into two groups based on the importance of WWW as a source of information and second using data driven segmentation on the basis of all other sources of information within each of the above groups. In the second step, Beritelli et al. (2007) used k-means cluster analysis with centroid method to form five clusters based on non-standardized items.

Also Boksberger and Laesser (2009) have used k-means cluster analysis to segment senior travel market by the means of travel motivations. To overcome unwanted homogeneity within a case they calculated a magnitude indicating the individual relative magnitude per item in relation to the overall mean of all items of travel motivation per case. As a results they found three clusters that all differed from each other in what they thought of as important or unimportant.

As the review of earlier market segmentation literature in tourism shows, a large number of different segmentation methods have been used and the results have varied from one study to another. However, it is clear that there is no consensus on if the data should be standardized or not. There are studies that have used non-standardized data and studies that have standardized data before calculating the segmentation solution. It is critical that the segments are not just results of a statistical analysis but that they exist in real-world and are practically usable (Pesonen, 2014).

In most studies, items used in the analysis are of categorical scales but treated as continuous. For example, a popular statistic packages SPSS has so called 2-step clustering that allows use both categorical and numeric variables. It also produces Information criteria (AIC, BIC), which are tools for comparing different clustering solutions and would be recommendable at least for comparative purpose (Martínez et al., 2006). However, this article concentrates only on K-means clustering due to its popularity in tourism studies.

Based on the literature review, the purpose of this paper is to examine the effect of response style when using k-means cluster analysis to segment tourists and demonstrate the effects of data standardization. This study contributes to the segmentation literature by criticizing one of the most used segmentation methods. Further-more, it is the first study in tourism to examine in detail response style effects in a sample based on a single nationality. The results of different segmentation methods are rarely compared and the current study contributes to the existing knowledge by comparing four different methods that use popular k-means clustering to find segments.

Data and methods

The data was collected on the Finnish Cottage Holidays website www.lomarengas.fi during summer 2009 using banner advertisement (Figure 1). Respondents were asked to rate the importance of 31 different motivation statements based on earlier literature (e.g. Kastenholz et al., 1999; Frochot, 2005; Komppula, 2005; Park & Yoon, 2009) using 7-point Likert scale (ranging from 1=not at all important to 7=very important). Data was analyzed using SPSS Statistics 17.0. Altogether 1043 questionnaires were completed by users of the website, of which 316 had to be deleted because of missing answers. Of the remaining 727 questionnaires 18 were deleted because standard deviations of motivation scores were 0. Altogether 709 suitable responses for the analysis methods of this study were used. Of the 31 motivation statements asked 25 were used for the analysis in this study. Six most skewed motivation statements were removed from analysis to increase the reliability of results.

There are multiple ways to check how strongly the data are affected by response styles (Dolnicar, 2006b). Most of these are suitable for multinational studies. As the focus of this study is the effect of response style in cluster analysis and getting clusters of people who said everything was important, some who said everything had little importance and some clusters in between, an approach suggested by Hair et al. (2010) is used. To remove ARS, the mean score across all motivation statements was calculated for each respondent. The mean score was used to calculate relative importance of each item for each respondent.

$$ARS = \frac{x - \bar{x}}{\bar{x}} \quad (1)$$

where x is the motivational score and $\bar{x}$ is mean of all motivation items of the respondent.

New motivation scores were used to calculate four different segmentation solutions similar to the first part of the results. Hair et al. (2010) call this method as within-case or row-centring standardization, which can be quite effective in removing response style effects and is especially suited to many forms of attitudinal data. According to Schaninger and Buss (1986), the rationale of a standardized cluster solution is that it produced more interpretable cluster solutions and demonstrated clearer differences in the relative determinance of different attributes within each cluster.

To remove ERS, approach suggested by Dolnicar and Grün (2009) was used. All motivation scores were divided by the individual standard deviation scores calculated earlier. For the last part of the analysis both ARS and ERS were removed by first subtracting of individual means and after that by division of individual standard deviations.

$$ERS = \frac{x}{s} \qquad (2)$$

where x is individual motivational score and s is standard deviation.

To cluster the data, a very popular method of k-means clustering was used. There are two different types of clustering procedures, hierarchical and nonhierarchical methods (Hair et al., 2010). In hierarchical procedures, observations are combined into hierarchy or a treelike structure whereas nonhierarchical clustering procedures assign objects into clusters once the number of clusters is specified. K-means algorithm belongs to nonhierarchical clustering algorithms and portions the data into k number of clusters specified by the researcher. K-means clustering algorithm maximizes the distance between clusters and minimises the distance of observations from one another in a cluster (Hair et al., 2010). The correct number of segments is also difficult to determine. In this study, different solutions based on different number of segments from two to six are compared. Typically earlier market segmentation studies in tourism have identified segmentation solutions within this range.

Data collection	1. Online questionnaire, www.lomarengas.fi users 2. 1043 completed questionnaires 3. 709 usable questionnaires for data analysis
Data standardization	1. Raw data, non-standardized data 2. ARS: Within-case standardization 3. ERS: Motivation scores divided by the individual standard deviation scores 4. Both ARS and ERS were removed by first subtracting individual means and then by division of individual standard deviations.
Data analysis	1. K-means cluster analysis, raw data 2. K-means cluster analysis, ARS standardized data 3. K-means cluster analysis, ERS standardized data 4. K-means cluster analysis, ARS+ERS standardized data

Figure 1. Research methodology

Results

Cluster analysis without removing response style effects

K-means cluster analysis was used to segment respondents based on their original scores for 25 motivation items. Four different cluster solutions were calculated to demonstrate how the number of clusters affects distribution of differences among segments and the segmentation results.

Table 1 depicts the clustering results of k-means cluster analysis based on non-standardized data. Results are presented by calculating how many times each segment has either the lowest mean or the highest mean among all segments when number of cluster increases from two to six. If two segments shared the same score, both were counted. From Table 1 it can be seen that using k-means analysis on original scores results in segments that differ from each other mostly by their response styles. Especially in two and four cluster solutions, it can be seen that clustering is based on extreme response styles and middle answering categories. Same can be seen also from other cluster solutions. Especially extreme response styles can be seen from all cluster solutions: there is always a segment with clearly higher average scores than other segments. In all solutions, except for the one with four clusters, there are also segments with clearly more second highest means and lowest means. In three cluster solution two different clusters, 3 and 4, share the number of lowest means. As can be seen from Table 1, using cluster analysis on original data fails to find distinctive segments and instead seems to clusters respondents according to their response styles.

Table 1. Five cluster solutions and distribution of means between clusters using original data.

	Two cluster solution		Three cluster solution		Four cluster solution		Five cluster solution		Six cluster solution	
	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means
Cluster 1	25	0	1	12	0	0	0	0	2	7
Cluster 2	0	25	24	0	23	0	0	17	0	17
Cluster 3			0	13	2	0	24	0	1	0
Cluster 4					0	25	1	0	0	1
Cluster 5							0	0	23	0
Cluster 6									0	0

Cluster analysis with acquiescence response styles effects removed

K-means cluster analysis was also conducted for 25 motivation items with scores standardized to remove acquiescence response style effects. The results are presented in Table 2. All clusters have items that they value more than other clusters, except for fifth cluster in six cluster solution. All clusters have also items that they value less than other clusters, except for first and third cluster in six cluster solution. This means that segments are distinctive from each other in many ways and segmentation results are practical.

Table 2. Five cluster solutions and distribution of means between clusters using ARS data.

	Two cluster solution		Three cluster solution		Four cluster solution		Five cluster solution		Six cluster solution	
	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means
Cluster 1	15	10	5	6	6	5	7	5	8	0
Cluster 2	10	15	5	11	2	2	3	2	10	2
Cluster 3			15	8	3	11	11	6	3	0
Cluster 4					14	7	1	3	2	3
Cluster 5							3	10	0	9
Cluster 6									2	11

Cluster analysis with extreme response styles removed

The results in Table 3 suggest that removing extreme response style effects does not increase the validity of the results. The results of k-means cluster analysis are very similar to results obtained using original, non-standardized data. There are clusters that have highest and lowest means in almost all motivation statements in all cluster solutions. Solution with six clusters has most distinctive segments but nevertheless the first cluster has highest mean in 20 motivation statements and third cluster has lowest mean score in 23 statements.

Table 3. Five cluster solutions and distribution of means between clusters using ERS data.

	Two cluster solution		Three cluster solution		Four cluster solution		Five cluster solution		Six cluster solution	
	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means
Cluster 1	1	24	0	24	0	0	0	23	20	0
Cluster 2	24	1	0	1	0	24	1	0	4	0
Cluster 3			25	0	25	0	24	0	0	23
Cluster 4					0	1	0	2	1	0
Cluster 5							0	0	0	2
Cluster 6									0	0

Cluster analysis with extreme response styles and acquiescence response styles removed

For the last part of the results both ARS and ERS are removed from the original data. The results presented in Table 4 demonstrate that this approach has produced very distinctive segments in all cluster solutions examined in this study. Only cluster 2 in five cluster solution and cluster 6 in six cluster solution do not think any motivations more important than other segments.

Table 4. Five cluster solutions and distribution of means between clusters using ARS and ERS data.

	Two cluster solution		Three cluster solution		Four cluster solution		Five cluster solution		Six cluster solution	
	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means	Number of highest means	Number of lowest means
Cluster 1	14	11	9	12	16	0	5	3	5	3
Cluster 2	11	14	13	1	3	11	0	2	2	7
Cluster 3			3	12	1	10	12	1	2	4
Cluster 4					5	4	3	10	14	1
Cluster 5							5	9	2	6
Cluster 6									0	4

Conclusion

The goal of this study was to examine and demonstrate how response style effects in a single nationality context affect cluster analysis results. For this purpose k-means cluster analysis with two to six cluster solutions were calculated for original, non-standardized data. Then three additional versions of the data set were created: one with acquiescence response styles removed, one with extreme response styles removed, and one with both response styles removed. Each of these data sets were analysed with k-means cluster analysis. The results are presented in Tables 1 to 4. The results in Table 1 are very similar to response style effects as described by for example Hair et al. (2010). Segments with high means and segments with low means were found in all different cluster solutions. However, when comparing results in Table 1 to those in Table 2 with ARS removed, it can be seen that results are completely different. Using row-centring standardization completely changes cluster analysis results. This suggests that even though the data consists only of Finnish respondents, response style effects play a great role in segmentation results. This means that just focusing on a single culture or nationality in a segmentation study does not mean that the effects of possible response styles should not be examined.

Removing extreme response styles by dividing original scores with individual standard deviation does not seem to have as significant effect on cluster analysis results as does removing acquiescence response styles, as can be seen from Table 3. This demonstrates that either ERS does not affect the data used in this study or that k-means cluster analysis is prone to find cluster with high scores in all items and low scores in all items when ERS is removed. However, removing both ARS and ERS from data seems to work well with k-means cluster analysis, but this can be mostly because of ARS, as can results in Table 2 and 4 are very similar. After response style effects were removed segments differed from each other in many ways. Almost all segments in Tables 2 and 4 had something they valued more than other segments and something they did not think of as important. This helps to identify what different segments really want. Instead of naming segments as Want-it-all tourists or Passive tourists, segments that regard some very particular motives or benefits are important or unimportant for them can be found. In this case, tourists are not segmented according to their response styles, thus serving marketing management purposes.

Discussion

It is not easy to determine what segmentation solutions are correct. It is even more difficult to determine what the best solution for the success of market segmentation scheme is. In a way, all 20 different segmentation solutions presented in this study are correct as only methods generally accepted in tourism segmentation literature are used. In data driven market segmentation, the correct solution is based on subjective assessment and determined by the researcher (Pesonen, 2014). However, as this study demonstrates, the possibilities are often almost limitless. The solution presented in research papers is often based on statistical methods and judgment and expertise of the researcher. The goal of the researcher is to find the solution that produces segments that differ from each other the most, represent the reality the best and are the most practical from a practitioner's viewpoint. Determining the best solution is out of the scope of this study, but nevertheless removing ARS produces most distinctive segments in this case. This approach has also been used and validated in studies by Beh and Bruyere (2007) and Beritelli et al. (2007).

According to Weijters (2006), response styles are affected by for example demographics and data collection methods. These factors also most likely influence the results of this study. However, this only denotes that different data are affected by different response styles. In this study, only ARS and ERS are examined, but other response styles can also affect the results of consumer studies meaning that data should always be carefully examined before analysis also in regard to response styles.

In data-driven tourism segmentation studies, several different scales have been used, most common being ordinal scale (Dolnicar 2002). Used scale has strong influence on response style effects. In this study, as in many earlier studies, seven-point Likert scale is used. Based on these results, it seems that Likert scale is quite vulnerable to response style effects. As an alternative, researchers and practitioners could use binary scales as it is much less likely to be affected by response styles. However, this results in a loss of some nuances in the data.

If the possible existence of response style effects in the data is not examined before doing segmentation analysis, there is a danger that the conclusions of a segmentation study are incorrect. For example the "Want-it-all" segment very common in earlier studies could be the result of an incorrect data analysis instead an actual segment existing in the market place. Authors who conduct market segmentation studies should report how response style effects have been examined in the data to demonstrate that response styles do not affect the results. There are many factors possibly affecting response styles besides culture, such as age, gender, and education (De Jong et al., 2008). This means that the existence of response styles have to be accounted for either when designing a questionnaire or when analysing the data.

This study demonstrates the effects of response styles when using cluster analysis and the results show that response styles determine in the formation of segments. In this study, only one segmentation method, k-means cluster analysis, is examined, but based on the results it can be argued that response styles can also have strong impact when other methods such as neural networks, hierarchical cluster analysis and factor-cluster segmentation are used. Even though response styles are examined in this study, it cannot be guaranteed that response styles are cleaned from the data despite data pre-processing as response styles per se are not measured. Only the possible effects of response styles on k-means cluster analysis in tourism segmentation are presented, demonstrating the changes in results.

This is the first study that particularly examines k-means cluster analysis in tourism and response style effects. As can be seen from the results of this study response styles have significant effect on segment formation also in single nation context. This is especially important for tourism industry because of its often multinational and multicultural nature. Response styles affect results even when all the respondents are from a single country, meanings that multinational and multicultural studies are almost surely affected by response styles. Especially interesting is to know what makes respondents use certain response styles. Even though response styles are well known factor in social sciences they have not been very often discussed in the context of tourism segmentation. This study aims to raise awareness of the effects of response styles and suggests that especially in segmentation studies response style effects should be examined before any additional analyses are conducted. In this study, response styles are examined using real world example. In future studies, simulated data could be used to study the topic, allowing researchers to measure response styles and best ways to remove them before segmenting tourists. Other clustering methods such as bagged clustering and two-step clustering should also be studied in regards to response style effects. It should be noted here that the data does not represent any population, but for the purpose of this paper it is not a major limitation as the goal is not to generalize the results to a population.

References

Alpert, M. I. (1980). Unresolved Issues in Identification of Determinant Attributes. In *Advances in Consumer Research VII* (Ed. Olse, J.C.), pp. 83–88. Association of Consumer Research, Ann Arbor.

Baumgartner, H., & Steenkamp J.-B. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research, 38*(2), 143–156.

Beh, A., & Bruyere, B. (2007). Segmentation by visitor motivation in three Kenyan national reserves. *Tourism Management, 28*(6), 1464–1471.

Beritelli, P., Bieger, T., & Laesser, C. (2007). The Impact of the Internet on Information Sources Portfolios. *Journal of Travel & Tourism Marketing, 22*(1), 63–80.

Bieger, T., & Laesser C. (2002). Market Segmentation by Motivation: The Case of Switzerland. *Journal of Travel Research, 41*(1), 68–76.

Boksberger, P. & Laesser, C. (2009). Segmentation of the senior travel market by the means of travel motivations. *Journal of Vacation Marketing, 15*(4), 311–322.

Calantone, R.J., & Sawyer, A.G. (1978). The Stability of Benefit Segments. *Journal of Marketing Research, 15*(3), 395-404.

Chun, K., Campbell, J.B. & Yoo J.H. (1974). Extreme response style in cross-cultural research. J*ournal of Cross-Cultural Psychology, 5*(4), 465–480.

Chung, K., Oh, S., Kim, S., & Han, S. (2004). Three representative market segmentation methodologies for hotel guest room customers. *Tourism Management, 25*(4), 429–441.

Clarke, I. (2001). Extreme response style in cross-cultural research. *International Marketing Review, 18*(3), 301–334.

De Jong, M. G., Steenkamp, J. B. E. M., Fox, J.P., & Baumgartner, H. (2008). Using Item Response Theory to Measure Extreme Response Style in Marketing Research: A Global Investigation. *Journal of Marketing Research, 45*(1), 104–115.

Dibb, S. (1998). Market segmentation: strategies for success. *Marketing Intelligence & Planning, 16*(7), 394–406.

Dolnicar, S. (2002). A Review of Data-Driven Market Segmentation in Tourism. *Journal of Travel and Tourism Marketing, 12*(1): 1–22.

Dolnicar, S. (2006a). Data-driven Market Segmentation in Tourism – Approaches, Changes Over Two Decades and Development Potential, *CD Proceedings of the 15th International Research Conference of the Council for Australian University Tourism and Hospitality Education (CAUTHE)*, Australia.

Dolnicar, S. (2006b). Are we drawing the right conclusions? The dangers of answer format effects in empirical tourism research. *Tourism Analysis, 11*(3), 199–209.

Dolnicar, S., & Grün, B. (2007). Cross-cultural differences in survey response patterns, *International Marketing Review, 24*(2), 127–143.

Dolnicar, S., & Grün, B. (2009). Response Style Contamination of Student Evaluation Data. *Journal of Marketing Education, 31*(2), 160–172.

Dolnicar, S., Grün, B. & Le, H. (2008). Cross-cultural comparisons of tourist satisfaction: Assessing analytical robustness. In: Yuksel, A. (Ed.), *Tourist Satisfaction and Complaining Behaviour: Measurement & Management Issues in the Tourism and Hospitality Industry*, 137–150, New York: Nova Publishing Sciences.

Everitt, B. S. (1993). Cluster Analysis. New York: Halsted Press.

Frochot, I. (2005). A benefit segmentation of tourists in rural areas: a Scottish perspective. *Tourism Management, 26*(3), 335–346.

Füller, J., & Matzler, K. (2008). Customer delight and market segmentation: An application of the three-factor theory of customer satisfaction on life style groups. *Tourism Management, 29*(1), 116–126.

Greenleaf, E. (1992). Measuring Extreme Response Style. *Public Opinion Quaterly, 56*(3), 328–351.

Hair, J., Black W., Babin B. & Anderson R. (2010). Multivariate data analysis. 7th edition. Upper Saddle River, New Jersey: Prentice Hall.

Hui, C.H., & Triandis, H.C. (1989). Effects of culture and response format on extreme responsestyle. *Journal of Cross-Cultural Psychology, 20*(3), 296–309.

Kastenholz, E., Davies, D. & Paul, G. (1999). Segmenting Tourism in Rural Areas: The Case of North and Central Portugal. *Journal of Travel Research, 37*(4), 353–363.

Komppula, R. (2005). Pursuing customer value in tourism – a rural tourism case-study. *Journal of Hospitality & Tourism, 3*(2), 83–104

Martínez, I.N.1., Morán, J.M., & Peña, F.J. (2006). Two-step cluster procedure after principal component analysis identifies sperm subpopulations in canine ejaculates and its relation to cryoresistance. *Journal of Andrology, 27*(4),596–603.

May, J., Bastian, C., Taylor, D., & Whipple, G. (2001). Market Segmentation of Wyoming Snowmobilers. *Journal of Travel Research, 39*(3), 292–299.

Middleton, V. T., & Clarke, J. (2001). Marketing in Travel and Tourism (3rd ed). Oxford, UK: Butterworth Heinemann.

Park, D.-B., & Yoon Y.-S. (2009). Segmentation by motivation in rural tourism: A Korean case study. *Tourism Management, 30*(1), 99–108.

Pesonen, J. (2014). Testing Segment Stability: Insights from a Rural Tourism Study. *Journal of Travel & Tourism Marketing, 31*(6), 697–711.

Rosmalen, J., van Herk, H., & Groenen, P. (2010) Identifying Response Styles: A Latent-Class Bilinear Multinomial Logit Model. *Journal of Marketing Research, 47*(1), 157–172.

Schaninger, C., & Buss, W.C. (1986). Removing Response-Style Effects in Attribute-Determinance Ratings to Identify Market Segments. *Journal of Business Research 14*(3), 237–252.

Weijters, B. (2006). Response Styles in Consumer Research. *Phd. dissertation submitted to the Faculty of Economics and Business Administration*, Ghent University. Accessed from: http://www.feb.ugent.be/nl/Ondz/Proefschriften/Weijters_B_proefschrift.pdf.