

# Testing segment stability: insights from a rural tourism study

This paper is author's draft of the paper published in:

[Testing Segment Stability: Insights from a Rural Tourism Study](#)

Juho Antti Pesonen

*Journal of Travel & Tourism Marketing*

Vol. 31, Iss. 6, 2014, pp. 697-711

Doi: 10.1080/10548408.2014.884965

## Abstract

Majority of earlier segmentation studies have focused on finding segmentation solutions rather than on applying the results to practice and testing the viability of the results. In this study a unique opportunity is used to test how the segmentation solutions of an earlier rural tourism segmentation study conducted in 2009 represent rural tourist segments in 2011 and how well rural tourists can relate to the segments found in the earlier study by using multiple choice questions. Also different segmentation methods are compared regarding their accuracy to identify the segments. These results show that the four segments identified in the earlier study continue to exist two years later as respondents are able to relate to the segments quite well. However, segment sizes are crucially different and there is some overlap between segments. Travel motivations measured using binary scale produce more accurate segments than if motivations were measured using Likert-type scale.

**Key words:** market segmentation, rural tourism, cluster analysis, segment stability, segmentation criteria

## Introduction

According to Dolnicar and Grün (2008, pp.63), ideal market segments "contain tourists with similar tourism needs and behaviors, similar sociodemographic profiles, who are profitable, who could easily be reached with marketing communication messages, who match the strengths of the tourism destination or business, and whose needs are not catered for by major competitors." Besides these criteria many authors (e.g. Dibb & Simkin, 2010; Morrison, 2002) regard that ideal segments should also be stable over certain periods of time. However, in the market segmentation literature and in the case of a posteriori segments tracking of market segment trends is neglected (Dolnicar, 2004). Longitudinal market segmentation studies are not very common and the topic of segment stability is very seldom discussed in the literature.

There are two basic ways to segment markets. In a priori or common sense segmentation (Dolnicar, 2002) individuals are grouped according to a criterion known in advance, such as age or gender. In a posteriori or data-driven segmentation (Mazanec, 2000) or post hoc segmentation (Wedel & Kamakura, 1998) an empirical data set is analyzed using quantitative techniques in order to derive a grouping. Most of the earlier market segmentation research in tourism has focused on finding a segmentation solution based on common sense or data-driven research and then validating those results by comparing external factors such as socio-demographics, activities or buying behaviors. According to Everitt (1993), most studies conducted a posteriori segmentation use a technique belonging to the family of cluster analysis. However, cluster analytic techniques will always render a result whether or not there really are

meaningful segments in the data (Dolnicar, 2003). This problem can partially be overcome by using several different techniques when analyzing the data but still the problem persists. The usefulness of any data-driven segment identification is dependent on two things: the quality of the data and the best possible use of cluster analysis or any other segmentation method (Dolnicar, 2002), making data collection and analysis a crucial part of the data-driven market segmentation process.

This study examines both data collection and data analysis in data-driven market segmentation in tourism. The purpose of this study is to test the results of an earlier segmentation study and to see how stable market segments identified in that study are and what are the different means to find them again, thus validating the results of the earlier study. By comparing different clustering methods this study also examines the accuracy of different solutions. The present study contributes to tourism segmentation literature by examining the stability of an earlier segmentation study using separately collected empirical data and measuring segmentation effectiveness as well as to choosing the correct segmentation solution. Cluster overlap in segmentation is also discussed.

The context of this study is rural tourism. In the tourism and hospitality literature market segmentation is a popular topic. Especially in rural tourism segmentation has proven to be an important field of study (e.g. Park & Yoon, 2009; Pesonen 2012). Tourism is often seen as an opportunity for rural economic development (Cawley & Gillmor, 2008). People come from urban centres to the countryside to enjoy the natural settings, peace, quiet and leisure activities it can offer (Finnish Ministry of Agriculture and Forestry, 2013). There has been ample interest in the literature in segmenting rural tourists. Park and Yoon (2009, pp. 99) state that “we need to know the causative factors and influences by which tourists in rural areas are motivated to become included in various market segments.” It is also important to gain a better understanding of the demand for rural tourism (Molera & Albaladejo, 2007) or rural tourist behavior (Pesonen, 2012), providing deeper insight into the profiles of rural tourists (Frochot 2005) and the lack of research addressing the rural tourism sector (Kastenholz et al. 1999).

## **Literature Review**

### **Market segmentation and segment quality criteria**

Common to many market segmentation studies in tourism is the use of clustering methodology to find a segmentation solution (e.g. Kastenholz et al. 1999; Molera & Albaladejo, 2007; Frochot, 2005; Park & Yoon, 2009; Pesonen, 2012). According to Dolnicar (2002), clustering has become a very popular way of identifying market segments based on survey data. However, the results of segmentation studies using the cluster partitioning method are more than questionable because of some very fundamental weaknesses (Dolnicar, 2002). Often in segmentation studies stability is tested and results are assumed to be valid and reliable if the segments are found repeatedly in the data set. As cluster is an exploratory tool (Dolnicar, 2002) the results constitute one of many possible solutions. It is important that solutions are useful for industry purposes, and according to Dibb and Simkin (2010), close attention has to be paid to segment quality.

In rural tourists have been segmented primarily by travel motivation (Park & Yoon, 2009) or by the benefits they seek (Kastenholz et al. 1999; Frochot, 2005; Molera & Albaladejo, 2007). A combination of the two has also been used (Pesonen, 2012). Oh and Schuett (2010) explored a visitor segmentation approach based on rural visitor spending behavior. Common to all these studies is the use of cluster analysis to identify the market segments.

Many studies have been presented assessing the various segment quality criteria (Dibb & Simkin, 2010). Kotler and Keller (2006) propose five key criteria for market segments to be useful: market segments have to be measurable, substantial, accessible, differentiable and actionable. Middleton et al. (2009) also use five criteria and state that segments have to be discrete, measurable, viable, appropriate and sustainable. Morrison (2002) also presented eight criteria for evaluating the success of a segmentation scheme. According to Morrison (2002) segments have to be homogenous, measurable, substantial, accessible, defensible, competitive, compatible, and durable. According to Wilkie (1994) there are three criteria for a true market segment. Members of the segment must be similar to other members of that segment as well as different from members of other segments. Members of a segment should also respond in a similar manner to a specific marketing mix. The organization in question should also be able to develop an efficient marketing mix for each segment.

Dibb and Simkin (2010) examined the practical application, impact and efficacy of segment quality criteria using a longitudinal case study from the Eastern European mobile phone market. They reviewed the earlier segmentation literature to identify the published segment quality criteria. As a result of a literature review they listed six segment quality themes and one underlying requirement for segmentation. The underlying requirement was that the market to be segmented must be heterogeneous. The six segment quality themes are homogenous segments, segment size and potential profitability, segment stability, segment accessibility, segment compatibility, and segment actionability.

Testing the quality of the segments is a crucial step in the market segmentation process (Dibb & Simkin 2010). Despite the importance of segment stability, it has not often been reported in the literature. According to Dibb and Simkin (2010), using qualitative criteria to evaluate the quality of the segments makes it possible to assess the intuitive managerial logic of the recommendation. Segmentation results can hardly be useful for managerial purposes if they are not stable over time, meaning that the results should be useful beyond the time of data collection. The segments should also be easy to interpret for companies aiming to utilize the results. It can be a limitation in a market segmentation study if people cannot relate to segments identified using statistical methods. In this study attention is paid especially to segment size and segment stability.

### **Identifying market segments**

Numerous methods have been proposed to find the segmentation solution from a quantitative data set. The most common of these is the use of cluster analysis, which refers to a large number of techniques for grouping respondents according to their similarities and differences (Dolnicar, 2003). Dolnicar (2003) also states that each technique is different and typically leads to different segmentation solutions. Therefore special attention should be paid to selecting the algorithm to be imposed on the data.

There is no clear consensus in the market segmentation literature whether ordinal, nominal or metric data should be used when measuring the segmentation base. According to the literature review by Dolnicar (2002), most earlier market segmentation studies in tourism have used ordinal data, which is used twice as often as nominal data. However, studies comparing segmentation results between ordinal and nominal data in the same data set are virtually nonexistent in tourism research. Studies comparing different segmentation approaches in the same study have been very uncommon in the tourism segmentation literature. There has been considerable debate over which bases and statistical approaches provide the best segmentation solutions (Moscardo et al. 2001).

A great deal of criticism has been leveled at cluster analysis in marketing and management literature. For example, Ketchen and Shook (1996) analyzed 45 published strategy studies and

found that the implementation of cluster analysis has been often less than ideal. Dolnicar and Leisch (2003) presented a bagged clustering approach to segment visitors in order to increase the stability of segmentation results. They suggest that further research is needed to prove that the segments extracted are not artifacts of the partitioning procedure. Dolnicar and Grün (2008) argued that when finding segmentation solutions the traditional "factor-cluster segmentation" is not generally the best procedure to identify homogenous market segments. They conclude that clustering the raw data directly is the superior alternative to "factor-cluster segmentation" as it identifies the true heterogeneity in the data better.

There is also a possibility that market segments can overlap each other. According to Saunders (1994) in conventional cluster analysis observations are separated into mutually exclusive subsets which are then described, whereas in reality segment members could belong to more than one group. Baloglu and Uysal (1996) found significant overlap with German overseas pleasure traveler segments using canonical correlation analysis. However, cluster analysis, the most popular segmentation method in tourist segmentation (Dolnicar, 2002), does not allow segment overlap and the topic of overlapping segments with means cluster analysis is not often discussed. In this study overlap of segments found using k-means cluster analysis is explored.

Pesonen (2012) conducted a market segmentation study on travel motivations among Finnish rural tourists in summer 2009. He identified four rural tourist segments: Wellbeing Tourists, Home Region Tourists, Family Tourists and Social Tourists. In this study the validity of the results of an earlier rural tourist segmentation study by Pesonen (2012) is examined using qualitative criteria, that is, how users of the same website can relate to segments found in the earlier study. The idea is quite similar to that applied in the study by Horneman et al. (2002), who asked respondents to rate the preference of discriminating factors of an earlier segmentation study to categorize senior travelers into those segments.

According to Dibb and Simkin (2010), one of the difficulties marketers face is testing the quality and robustness of segments. Saunders (1994) states that statistically significant results have to be usable and accepted by managers. Managers have to believe in the clusters formed, recognize them, and see how they can be used. Market segments cannot be only academic concepts; their effectiveness must be proven in reality. By examining how well tourists can relate to segments identified using statistical methods it can be shown that the segments also exist in reality.

It is recognized in the literature that individuals belong to different market segments during different life phases. According to lifecycle theory, as people get older their needs and social roles change, and likewise their travel motivations (Boksberger & Laesser, 2009; Horna, 1994). Thus the important question in segmentation is not if a person belongs to one segment or another. The question is if the segments exist at any given time and behave and respond to marketing mix in the way they are expected to. It is also important that the tourists in the same segment behave in the same way. The lifecycle theory from the segmentation point of view means that even though a person may change the segment of which he or she is a member, that person will still behave as expected when included as a member of a certain segment.

The use of Likert scales and cluster analysis has also been criticized. Despite being a much used tool for market segmentation, cluster analysis has several pitfalls. One of these is that it always produces a solution regardless of the data or the if there are or are not patterns in the data that can be regarded as segments (Dolnicar, 2003). Cluster analysis is also highly susceptible to response style effects, especially when using data measured using Likert-type scales. According to Hair et al. (2010), there is a possibility that when clustering data is collected using, for example, a number of ratings on a 10-point scale, we could end up with clusters of people who said everything was important, some who said everything was of little importance and maybe some clusters in between. This is called response-style effect and results resembling it can be

seen in many different segmentation studies in tourism (see e.g. Bieger & Laesser, 2002; Chung et al., 2004; Füller & Matzler, 2008; Park & Yoon, 2009). Data driven segmentation is heavily dependent on the segmentation method and data analysis conducted. These, however, are very rarely discussed in the literature. There is a need to test different methods and thus improve the quality of data analysis when segmenting tourism markets.

Arimond and Elfessi (2001) demonstrated that multistate categorical survey data can be successfully used to cluster tourists. They stated that in the future other methods for clustering market segments with qualitative, categorical data should be investigated. Arimon and Elfessi (2001) as well as Green and Krieger (1995) suggest that it would be useful to run comparative studies using the traditional quantitative ratio scale segmentation methods.

In light of the literature review and the gaps found in the way market segmentation is conducted in the field of tourism and hospitality research, the purpose of this study is to examine the quality of the market segmentation results of an earlier study, in this case a rural tourism segmentation study based on travel motivations (Pesonen, 2012). This study examines rural tourists' perceptions of the clusters that should represent them and tries to find clusters similar to those reported by Pesonen (2012) by collecting new quantitative data and comparing the results to those reported by Pesonen (2012). This study has thus four research questions:

- 1) How well can Finnish rural tourists relate to previously identified rural tourist segments?
- 2) How much overlap there is between segments?
- 3) How accurately can market segments be rediscovered from new data using quantitative clustering methods?
- 4) How do travel motivation segments identified using Likert scale data differ from segments identified using binary data regarding the accuracy of the segmentation solution with self-selective segment membership?

## **Measurement**

The starting point for this study is the study conducted by Pesonen (2012) in 2009. In earlier market segmentation studies focusing on rural tourists in Finland four different rural tourist segments have been identified (Pesonen, 2012; Pesonen & Komppula, 2010). Pesonen (2012) collected a quantitative data during summer 2009 on the largest Finnish rural tourism website using a survey questionnaire. The survey focused on what motivates rural tourists to travel and what kind of destination attributes they value. Altogether 727 usable answers were collected by Pesonen (2012). Pesonen segmented the website users into four segments on the basis of their travel motivations using K-means cluster analysis with data preprocessed to account for response style effects by calculating average mean scores across all motivation statements for each respondent and using these scores to calculate relative importance of each item for each respondent. The segments identified were Social Travelers (N=213, 29.3 %), Wellbeing Travelers (N=164, 22.6 %), Home Region Travelers (N=148, 20.4%) and Family Travelers (N=202, 27.8 %) (Pesonen, 2012).

To examine segment stability over time and to compare different market segmentation methodologies three different kinds of data were collected in this study:

- 1) Travel motivations measured with 7-point Likert-type scale,
- 2) Travel motivations measured with binary scale (important / not important) and
- 3) What segment identified by Pesonen (2012) best describes the respondent in his or her own opinion.

Pesonen (2012) used 31 travel motivation statements, measured using 7- point Likert-type scale ranging from 1 (not at all important) to 7 (very important). To keep this questionnaire short and to get more responses only those travel motivations that differentiated the clusters the most were included. Including more items from the study by Pesonen (2012) would have made this questionnaire considerably longer and would not have helped in separating the clusters. Similar 7-point Likert-type scale that Pesonen (2012) used was utilized. The respondents were asked to rate the importance of the following 12 travel motivations:

- I would have a hassle-free vacation
- I would like to escape from a busy everyday life
- There would be an opportunity to be together as a family
- I could visit places my family comes from
- I would have a feeling like I was being pampered
- I would have an opportunity to be physically active
- I would have a "once in a lifetime" experience
- I would like to relax away from the ordinary
- I would have some control over the way things turn out
- I would experience different culture
- I would have a feeling of romance
- I would have a chance to meet interesting people.

To measure travel motivations on a binary scale travel motivation statements from the study by Bieger and Laesser (2002) were used. Altogether 10 travel motivations were used. Respondents were asked to choose at least one and at most three different travel motivations that were most important for them. This approach was used to make sure respondents had to think about their choices instead of just selecting everything that they thought would be nice during their holiday. These ten statements are general travel motivations found in many other tourism studies (abbreviations in parentheses):

- Participating in nightlife (nightlife)
- Enjoying comfort, spoiling myself (comfort)
- Taking and having time for my partner (partner)
- Taking and having time for my family (family)
- Enjoying landscape and nature (nature)
- Broadening my mind, enjoying sightseeing (culture)
- Being able to make flexible and spontaneous decisions (liberty)
- Doing something for my looks and well-being (body)
- Sports activities (sports)
- Enjoying the sun and water (sun).

To find out how well the respondents could relate themselves to the segments identified in the study by Pesonen (2012) descriptions of the four segments were presented (Table 1). The differences between segments reported by Pesonen (2012) were used to describe the four segments of rural tourists. Respondents were asked to choose the segment which best described them. In contrast to the study by Horneman et al. (2002), a respondent could either belong to the segment or not, making segment membership a binary score, whereas Horneman et al. (2002) used a 4-point Likert-type scale.

INSERT TABLE 1 HERE

TABLE 1. Segment descriptions

Segment name	Description
Social Tourists	During your holiday you want to meet interesting people, be active and maybe even have a feeling of romance. You appreciate a destination that has a rich history and culture and where you can meet new people outside your own family. You also like to have control over your own holiday. You prefer traveling with friends.
Family Tourists	You enjoy traveling with your family. Having fun, being together with your family, new "once in a lifetime" experiences and everything that is new and exciting motivate you to travel. You want the destination to be safe for the family and for your children to enjoy the destination. You also appreciate environmental friendliness in a destination.
Wellbeing Tourists	You want to escape from your busy everyday life to the peace and quiet of the countryside. You want to relax from the routine and hassle of the cities and enjoy privacy and comfort. You appreciate having no timetables , a peaceful atmosphere, good opportunities for outdoor activities and beautiful landscapes.
Home Region Tourists	You are interested in traveling to the region your family comes from.
Something else, what?	If none of the above options describes you as a rural tourist you can describe your own rural tourism behavior here.

Respondents were also able to choose several different segments. The purpose of this was to see how well a description of a single segment could be used to describe a rural tourist or if a combination of several segments would serve better.

### Data Collection

For the purposes of this study a new questionnaire was used to validate the segmentation results of the earlier study. The questionnaire was promoted to the users on the same website that was used by Pesonen (2012) as well as two other rural tourism websites. Pesonen (2012) argues that the four segments can be found among Finnish rural tourists that use the Internet so two additional web sites were included to test this. A large majority of responses, more than 90 per cent, came from the website [www.lomarengas.fi](http://www.lomarengas.fi) which was the website also used by Pesonen (2012). Responses from the three websites were compared regarding travel motivations and only small differences ( $p < 0.05$ ) were found (Table 2). Largest difference in is in the opportunity to be physically active which was considerably more important for users of website 3 compared to users from other websites. Website 3 rents accommodation near a skiing center which explains the difference. Eta squared values that determine the strength of the relationship between the variables (Moscordo et al., 2001) are very low, meaning that the website used to collect the responses does not have large effect on how respondents answered to the questionnaire regarding travel motivations.

TABLE 2. Differences in travel motivations between the three websites

Item	Lomarengas.fi	Website 2	Website 3	Sig. (2-tailed)	Eta squared
I would have a hassle-free vacation	5.63	5.48	5.70	p=0.504	0.001
I would like to escape from a busy everyday life	5.75	5.60	6.16	p=0.062	0.003
There would be an opportunity to be together as a family	5.64	5.60	5.59	p=0.945	0.000
I could visit places my family comes from	3.55	3.12	3.08	p=0.022	0.004
I would have a feeling like I was being pampered	4.81	4.63	4.97	p=0.383	0.001
I would have an opportunity to be physically active	4.64	5.01	5.54	p<0.001	0.011
I would have a "once in a lifetime" experience	4.18	4.31	4.41	p=0.549	0.001
I would like to relax away from the ordinary	6.24	6.39	6.57	p=0.038	0.004
I would have some control over the way things turn out	3.92	3.51	3.86	p=0.026	0.004
I would experience different culture	4.31	4.75	4.65	p=0.009	0.005
I would have a feeling of romance	4.70	4.38	4.95	p=0.091	0.003
I would have a chance to meet interesting people	4.39	4.43	5.05	p=0.088	0.003

A banner advertisement with a prize of a 400-Euro gift voucher was used to attract the users to the questionnaire. Data was collected in summer 2011, from the beginning of March to the end of August. Whereas Pesonen (2012) collected data only during summer season a longer time span was chosen for this study in order to increase the sample size. In order to examine if the month of visiting the websites affects the responses the 12 travel motivations measured with 7-point Likert-scale were compared by dividing respondents into groups based on which month they had answered to the questionnaire. Only some very small statistical differences were found in variables "I would have a hassle-free vacation" ( $F=2.794$ ,  $p=0.016$ ), "I would like to escape from a busy everyday life" ( $F=2.799$ ,  $p=0.016$ ), "There would be an opportunity to be together as a family" ( $F=2.387$ ,  $p=0.036$ ) and "I would like to relax away from the ordinary" ( $F=3.667$ ,  $p=0.003$ ). Relaxation, being with family and escape from a busy everyday life seem to be more important in March and in July than during other months whereas hassle-free vacation is especially important in July. Based on these results it can be argued that the time when respondent answered to the questionnaire does not have huge effect on the results of this study. Altogether 1937 responses were obtained, all from Finnish users. Of these 1772 were usable for this study. Most of the excluded responses were a result of duplicate answers from the same respondents, who had not answered regarding the importance of travel motivations or because they had answered all Likert scale questions with the same answer.

## **Methodology**

Original data from the year 2009 that Pesonen (2012) used was acquired from the author in order to compare the results between these two studies. The data is analyzed in five phases:

- 1) A sample profile is presented.
- 2) Importance of travel motivations between 2009 and 2011 are compared.
- 3) The accuracy of market segmentation results from an earlier study (Pesonen, 2012) are analyzed by asking rural tourists which market segment describes them the best.
- 4) Five different statistical segmentation methods are used to ascertain which can produce segments that best represent the four rural tourist segments identified by Pesonen (2012) .
- 5) The accuracy of statistical segmentation methods is scrutinized by comparing the results with the segments that rural tourists think describe them the best.

First, sample profile is presented and differences between 2009 and 2011 studies are compared using chi-square test and one-way analysis of variance (ANOVA). Then travel motivations measured using Likert-type scale were compared between data collected for the present study and data used by Pesonen (2012). Comparative analysis was conducted using independent samples t-test. All responses containing missing values in travel motivations were excluded from this analysis.

For the third part of the study the results from the question on how well people could relate to the segments are considered as the correct segments existing in the data. As cluster analysis does not support overlapping clusters only people relating to just one segment are included in the data analysis. This results in 1509 usable questionnaires. Then these segments are sought from the data using the most used data analysis methods in the market segmentation literature on tourism.

The chosen segmentation methods are based on the study by Dolnicar (2002). In her review of data-driven market segmentation in tourism she found that 45 per cent of studies used factor analysis to preprocess data before clustering and 44 per cent of studies using hierarchical algorithms used Ward's method to derive groupings. In this study principal component analysis with varimax rotation is used to preprocess data for the use of K-means cluster analysis as one possibility for data analysis. Varimax rotation was chosen as it is one of the most common rotations used in segmenting tourism markets (see e.g. Konu et al. 2011). Pesonen (2012) preprocessed the data by standardization before clustering to eliminate the effects of response styles. Last cluster method is to use only K-means without any data preprocessing, an approach proposed by Dolnicar and Grün (2008). Four different methods with a number of cluster solutions ranging from three to five are used on the data measured with a Likert scale. Cluster solutions from three to five are most common in tourism segmentation studies and the earlier study by Pesonen (2012) also identified four clusters. To analyze binary data a hierarchical clustering algorithm was used with squared Euclidean distances and Ward's method. Cluster memberships from three to five clusters were saved. All the results of these analysis can be found from the Internet in the address <http://goo.gl/lb99M>.

In the last part of the data analysis the accuracy of segmentation methods that most successfully identified the four aforementioned segments from the data was examined. Cross tabulations with chi square analysis were used to compare the accuracy of statistical segmentation methods by cross tabulating statistical segment membership and self evaluated segment membership.

## Results

There are some differences between respondents in age and gender. In the study conducted in 2009, 17 per cent of respondents were male. In the present study 26.3 per cent are male. The difference is statistically significant ( $\chi^2=24.35$ ,  $p<0.001$ ). The respondents are also younger in this study with an average age of 39 years compared to 43 years in 2009 study ( $F=40.37$ ,  $p<0.001$ ). However, Bieger and Laesser (2002) for example noted that the sociodemographic situation of individual travelers seem to be least relevant for motivation segmentation, meaning that in this regard differences in sociodemographics between the two samples should not decrease the validity of these results too much. Also Johns and Gyimóthy (2002) state that age, gender and income are only indirectly related to what a person will buy.

There are some differences in samples between the 2009 and 2011 studies regarding socio-demographic factors and travel motivations. Altogether 12 travel motivations were measured in this study as well as by Pesonen (2012). There are statistically significant differences in many travel motivations, measured using independent samples t-test (Table 3). All the motivations that differ have higher mean score in the study conducted during summer 2009 except for visiting places where respondent's family comes from. However, in both studies relaxing away from the routine was the most important travel motivation, followed by escape from busy everyday life and taking a hassle-free vacation.

TABLE 3. Sample differences in importance of travel motivations between 2009 and 2011 studies

Item	2009 mean	2011 mean	t (Equal variances assumed)	Sig. (2-tailed)
I would have a hassle-free vacation	<b>6.02</b>	5.62	6.118	$p<0.001$
I would like to escape from a busy everyday life	<b>6.03</b>	5.75	4.978	$p<0.001$
There would be an opportunity to be together as a family	<b>5.89</b>	5.64	3.707	$p<0.001$
I could visit places my family comes from	3.27	<b>3.51</b>	-2.733	$p<0.001$
I would have a feeling like I was being pampered	<b>4.91</b>	4.80	1.419	$p=0.006$
I would have an opportunity to be physically active	4.70	4.69	0.248	$p=0.156$
I would have a "once in a lifetime" experience	4.43	4.20	3.108	$p=0.804$
I would like to relax away from the ordinary	<b>6.42</b>	6.26	3.919	$p<0.001$
I would have some control over the way things turn out	<b>4.20</b>	3.89	4.172	$p<0.001$
I would experience different culture	<b>4.54</b>	4.35	2.564	$p=0.010$
I would have a feeling of romance	2.64	2.69	-0.681	$p=0.496$
I would have a chance to meet interesting people	4.40	4.41	-0.089	$p=0.929$

The results of how well rural tourists can relate to the segments found earlier are presented in Table 4. As can be seen, most of the respondents could relate to one of the segments presented to them, while 10.3 per cent would position themselves to two different segments, 1.2 per cent to three and nine respondents (0.5 %) think that all the segment descriptions fit them. Respondents reporting that none of the segments could describe them amounted to 2.9 per cent. Family Tourists was the largest segment with 37.1 per cent of respondents, closely followed by Wellbeing Tourists (34.1 %). Approximately ten per cent of respondents think that Social Tourists describes them best. Only four per cent found description of Home Region Tourists to fit them best.

INSERT TABLE 4 HERE

TABLE 4. Rural tourists relating themselves to segments

Segment name	Rural tourists relating themselves to segments
Social Tourists	177 (10.0 %)
Family Tourists	657 (37.1 %)
Wellbeing Tourists	605 (34.1 %)
Home Region tourists	70 (4.0 %)
People belonging to two segments	182 (10.3 %)
People belonging to three segments	21 (1.2 %)
People belonging to four segments	9 (0.5 %)
Something else	51 (2.9%)

Table 5 presents the results of the cluster analysis. All the details of data analysis can be found from the Internet in the address <http://goo.gl/lb99M>. First a principal component analysis (PCA) with varimax rotation was used on the twelve travel motivation statements. This resulted in three principal components explaining 51.7 per cent of variance. Cronbach alphas for the principal components were 0.745, 0.685 and 0.362. It should be noted that 0.362 is very low value for alpha but can be explained by the fact that only two items were included in the third principal component. Adding more items would have increased the reliability of PCA and variance explained but on the other hand increased the effort of the respondents. According to Dolnicar and Grün (2008), PCA with cluster analysis is an outdated method that does not provide enough benefits to justify its use. However, it is still widely used and was included also in this study to demonstrate its effects in identifying segments.

The regressions score from the principal component analysis was used in K-means cluster analysis and the segments were compared using the original Likert scale scores. K-means segmentation based on regression score from PCA managed to produce quite distinctive segments and as the number of clusters increased so did the distinctiveness of segments. In three cluster solution Family Tourists and Wellbeing Tourists segments were identified The

wellbeing tourists segment valued a hassle-free vacation, escape from everyday life and relaxation more than respondents in other segments. The other main segment was one containing respondents who valued all but the aforementioned three travel motivations more than respondents in other segments, making it a combination of home region travelers, wellbeing tourists and family tourists. In four and five cluster solutions Wellbeing Segment was easy to find. Those who regarded the feeling of romance as more important than other segments were regarded to be Social Tourist in all the cluster solutions in all data analysis methods.

Segments found using only K-means cluster analysis were very similar to those found using hierarchical cluster analysis with Ward's method and squared Euclidean distances. The wellbeing segment was again distinguishable in every solution as was the segment that valued feeling of romance more than other segments. However, in this case the wellbeing segment valued not only motivations related to relaxation but also all other motivations except for feeling of romance more than other segments.

Examining travel motivations measured with binary data provides better results. In all the solutions the three segments identified in the earlier study can be found. The reason why home region travelers are not to be found is because questions relating to that travel motivation were not measured in the study by Bieger and Laesser (2002). In the four-cluster solution family travelers divides into two segments, both very much motivated by being together with family and in the five-cluster solution the wellbeing segment is also divided into two. Otherwise the cluster membership is very stable, as can be seen from segment sizes.

INSERT TABLE 5 HERE

TABLE 5. Rural tourists relating to segments with different segmentation methods (N=1509)

	Can the segment be found?			
	Social Tourists (N=177)	Family Tourists (N=657)	Wellbeing Tourists (N=605)	Home Region Tourists (N=70)
Principal component analysis and K-means cluster				
3	No	Yes (703)	Yes (670)	No
4	Yes (419)	Yes (565)	Yes (475)	No
5	Yes (210)	Yes (530)	Yes (443)	No
K-means cluster				
3	Yes (554)	No	Yes (555)	No
4	Yes (503)	No	Yes (489)	No
5	Yes (323)	No	Yes (360)	No
Hierarchical cluster with Ward's method				
3	Yes (513)	No	Yes (415)	No
4	Yes (513)	No	Yes (415)	No
5	Yes (513)	No	Yes (415)	No
RSE preprocessing and K-means cluster				

3	Yes (646)	No	Yes (532)	No
4	Yes (524)	No	Yes (398)	Yes (446)
5	Yes (253)	Yes (389)	Yes (361)	No
Binary data with hierarchical cluster analysis				
3	Yes (379)	Yes (524)	Yes (606)	No
4	Yes (379)	Yes (266+258)	Yes (606)	No
5	Yes (379)	Yes (266+258)	Yes (392+214)	No

In this study also the accuracy of the segmentation solution compared to the self-selective segment membership was examined. Because K-means cluster analysis and hierarchical cluster analysis with Ward's method did not produce very distinctive segments their accuracy is not compared.

The results of the accuracy analysis are presented in Table 6. As can be seen from the table, principal component analysis and K-means cluster analysis based on the regression scores of PCA correctly classifies from 27.8 per cent to 35.5 per cent of respondents.

In the three cluster solution with K-means cluster analysis of RSE-standardized data 66.1 per cent of those regarding themselves as Social Tourists belonged to the Social Tourists segment found. In the three-cluster solution a total of 20.7 per cent, in four-cluster solution 17.4 per cent and in five-cluster solution 23.5 per cent of respondents were correctly classified, meaning that K-means cluster analysis based on regression scores from PCA outperformed cluster analysis based on RSE standardized data.

However, the cluster analyses based on Likert scale data are clearly inferior to those identified using binary data. The results are quite impressive, with more than half of respondents correctly classified into data driven segments. The results between three, four and five cluster solutions are the same as increasing the number of clusters only divided family and wellbeing tourists into smaller segments, still mainly motivated by being together with family or motivations traditionally connected with wellbeing.

INSERT TABLE 6 HERE

TABLE 6. Examination of segmentation accuracy

	Percentage of correctly classified segment members				Total (N=1509)
	Social Tourists (N=177)	Family Tourists (N=657)	Wellbeing Tourists (N=605)	Home Region Tourists (N=70)	
Principal component analysis and K-means cluster					

3	0 %	259 (39.4 %)	276 (45.6 %)	0 %	535 (35.5%)
4	67 (37.9%)	220 (33.5%)	172 (28.4%)	0 %	459 (30.4%)
5	49 (27.7%)	211 (32.1%)	159 (26.3%)	0 %	419 (27.8%)
RSE preprocessing and K-means cluster					
3	117 (66.1%)	0 %	188 (31.1%)	0 %	305 (20.2 %)
4	110 (62.1%)	0 %	153 (25.3%)	0 %	263 (17.4%)
5	58 (32.8%)	160 (24.4%)	137 (22.6%)	0 %	355 (23.5%)
Binary data with hierarchical cluster analysis					
3	69 (39.0 %)	400 (60.9 %)	368 (60.8 %)	0 %	837 (55.5 %)
4	69 (39.0 %)	194 + 206 (60.9 %)	368 (60.8 %)	0 %	837 (55.5 %)
5	69 (39.0 %)	194 + 206 (60.9 %)	253 + 155 (60.8 %)	0 %	837 (55.5 %)

## Discussion and Conclusion

The aim of this study was to examine the quality of segmentation results from an earlier study (Pesonen, 2012). This was done by analyzing how accurately rural tourists could place themselves into the segments identified by Pesonen (2012) and also by using data driven segmentation to find the market segments again. The rural tourists in this study could relate to the segments identified by Pesonen (2012). In light of the results it can be argued that the rural tourist segments proposed by Pesonen (2012) depict Finnish rural tourists fairly accurately. Most of the website users could relate to one particular segment. The largest segment is Family Tourists (37.1 %) closely followed by Wellbeing Tourists (34.1%). Home Region Tourists was a very small segment; only four per cent of respondents reported travel to the place of their family's origin as their sole motivation.

Only 2.9 per cent of respondents could not find themselves in any of the segments or in combinations of these. Some respondents (12.0%) chose at least two different segments to describe them best. Even though there is some overlap between segments, most of the respondents could categorize themselves as members of a single segment, making it easy to use the results. Saunders (1994) regard overlapping clusters as an unnecessary complication and suggest that researchers should aim to find easily interpretable clusters. Market segments have to be designed so that a single person cannot be a member of several segments at any given time. It is possible for a person to change from one segment to another but that also means

that he or she has to behave the same way as the other members in the same segment. This has always been a strength of common sense segmentation, as a person cannot belong to more than one segment at a time. This topic has not often been discussed in relation to data driven segmentation studies. This study suggests that segment overlap is not a major problem with the use of cluster analysis and the respondents were generally able to place themselves in only one travel motivation segment.

When comparing these results to those of Pesonen (2012) it can be seen that there is a great difference in segment sizes between these two studies. Whereas Pesonen (2012) found that the four segments had almost equal numbers of rural tourists, the results of the present study show that Family Tourists and Wellbeing Tourists are actually much larger segments than Home Region Tourists or Social Tourists when considering how well rural tourists relate to the segments. This study also attributes these differences between two different studies to the way the data was analyzed. The K-means cluster analysis used in this study produced only segments that are relatively equal in size. However, this is not always the case in real life.

When segmenting the data using binary measured motivations the results are much closer to the results of how people relate to segments. Even though motivation to travel to the region the respondent's family comes from was not measured with binary travel motivations, all the other segments were quite easily distinguishable. The segment sizes are also much more closer to the way people relate to different segments.

Segmentation based on binary travel motivations could also classify respondents much better to the segments regarded as correct, that is, the segments that people could relate to. More than half of the respondents belonged to the correct segment when the binary data was analyzed with hierarchical cluster analysis, whereas the data analysis based on 7-point Likert scale data could classify little more than 20 per cent of cases correctly at best.

In earlier studies a Want-it-all segment has been a very common rural tourist segment (Kastenholz et al. 1999; Park & Yoon, 2009). However, this segment is typical with cluster analysis. In this study too a Want-it-all segment was found when using K-means cluster analysis either on data based on principal component analysis or on unstandardized Likert scale scores. A Want-it-all segment was also found with hierarchical cluster analysis of unstandardized data. However, from a managerial view perspective a Want-it-all segment is very difficult to operate with. Warnings about this have also appeared in the literature (Hair et al. 2010).

This study supports the notion of using binary data to collect information for market segmentation purposes. This way some of the common problems, such as response style effects with Likert scale scores and segments of equal sizes with the use of K-means cluster analysis, can be avoided or alleviated.

Stability of segments over time is also very important for managerial purposes. The results are hardly useful for managers if they can be used only at the time of data collection and not in the future. The data for this study was collected on rural tourism websites and only from online respondents. In order, for example, to differentiate the offering on a website it is important for managers to have a stable solution and ways to examine change in segments over time. According to Hoek et al. (1996), including variables that predict consumer behavior to the segmentation base is very important in order to find segments that are fairly constant in nature. The segment descriptions of Pesonen (2012) and the present study are also based on travel companions. Social Tourists travel with friends, Family Tourists with family and Wellbeing Tourists with their partners. These segments can thus be regarded as quite stable over time.

The results also show that data-driven solutions from statistical segmentation methods, in this case K-means cluster analysis, should be interpreted with caution. Even though a solution may seem logical and correct, it does not necessarily represent the real world. Relative segment size

is an important factor in deciding which segments to target and the results of this study suggest that segmentation using K-means cluster analysis is only a starting point for market segmentation and that segments need to be carefully evaluated.

### **Limitations and future research**

There are some limitations to this study. Even though the sample of this study is similar to the sample used by Pesonen (2012), they are not identical. Even though both sets of data were collected from the same rural tourism website only two years apart there are many differences in the travel motivations, age, and gender of the respondents. It is unknown whether these differences are a result of changes in the sample or changes in the population between studies. However, the typical respondent can be described the same in both studies: a middle aged female who is mostly motivated to travel for relaxation, escape and a hassle-free vacation.

Want-it-all segments and Passive Tourist segments should be explored further. As in many earlier studies, these two segments were also found in this study with certain partitioning methods. In this study it was assumed that the four segments identified by Pesonen (2012) were the correct segments to be analyzed but it is unknown how the results might differ if people were asked to relate to segments that include Want-it-all Tourists and Passive Tourists.

The validation of the segments found using the categorical data is beyond the scope of this paper, but is a crucial next step in examining the usefulness of the segments presented in this study. The wording of the question on how well customers can relate to segments should be explored further. The topic of applying segmentation results to practice is very important (Dibb & Simkin 2010) and in tourism this field also requires further research.

The strength of quantitative studies lies in the generalizability of their results. However, despite extensive research on rural tourism segmentation the results are always destination or country specific. An alternative approach could be to use a combination of qualitative and quantitative segmentation. For example, Mackellar (2009) used qualitative methodology to segment festival participants on the basis of their behavior. A quantitative approach could be used to generalize the segmentation results of a qualitative study using, for example, the methodology presented in this study, but this requires further research.

## References

- Arimond, G., & Elfessi, A. (2001). A Clustering Method for Categorical Data in Tourism Market Segmentation Research. *Journal of Travel Research*, 39(4), 391–397.
- Baloglu, S., & Uysal, M. (1996). Market segments of push and pull motivations: a canonical correlation approach. *International Journal of Contemporary Hospitality Management*, 8(3), 32–38.
- 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.
- Cawley, M., & Gillmor, D., (2008). Integrated Rural Tourism: Concepts and Practice. *Annals of Tourism Research* 35(2), 316-337.
- 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.
- Dolnicar, S. (2002). A Review of Data-Driven Market Segmentation in Tourism. *Journal of Travel and Tourism Marketing*, 12(1), 1–22.
- Dolnicar, S. (2003). Using cluster analysis for market segmentation - typical misconceptions, established methodological weaknesses and some recommendations for improvement. *Australasian Journal of Market Research*, 11(2), 5–12.
- Dolnicar, S. (2004). Tracking data-driven market segments. *Tourism Analysis*, 8(2-4), 227-232.
- Dolnicar, S., & Leisch, F. (2003). Winter Tourist Segments in Austria: Identifying Stable Vacation Styles Using Bagged Clustering Techniques. *Journal of Travel Research*, 41(3), 281–292.
- Dolnicar, S., & Grün, B. (2008). Challenging "Factor Cluster Segmentation". *Journal of Travel Research*, 47(1), 63–71.
- Dibb, S., & Simkin, L. (2010). Judging the quality of customer segments: segmentation effectiveness. *Journal of Strategic Marketing*, 18(2), 113-131.
- Everitt, B. S. (1993). *Cluster Analysis*. New York, NY: Halsted Press.
- Finnish Ministry of Agriculture and Forestry, 2013. Preserving the vitality of the Finnish countryside. [http://www.mmm.fi/en/index/frontpage/rural\\_development.html](http://www.mmm.fi/en/index/frontpage/rural_development.html) (accessed 18.3.2013).
- 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.
- Green, P. E., & Krieger, A. M. (1995). Alternative Approaches to Cluster- Based Market Segmentation. *Journal of the Market Research Society*, 37(3): 2–11.
- Hair, J., Black W., Babin B., & Anderson, R. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, New Jersey: Prentice Hall.
- Hoek, J., Gendall, P., & Esslemont G. (1996). Market segmentation. A search for the holy grail? *Journal of marketing practice*, 2(1), 25–34.
- Horna, J. (1994). *The Study of Leisure – An Introduction*. Toronto: Oxford University Press.
- Horneman, L., Carter, R., Wei, S., & Ruys, H. (2002). Profiling the Senior Traveler: An Australian Perspective. *Journal of Travel Research*, 41(1), 23–37.

- Johns, N., & Gyimóthy, S. (2002). Market Segmentation and the Prediction of Tourist Behaviour: The Case of Bornholm, Denmark. *Journal of Travel Research*, 40(3), 316-327.
- Kastenholz, E., Davis D., & Paul G. (1999). Segmenting tourism in rural areas: the case of North and Central Portugal. *Journal of Travel Research*, 37(4), 353-363.
- Ketchen, D., & Shook, C. (1996). The Application of Cluster Analysis in Strategic Management Research: an Analysis and Critique. *Strategic Management Journal*, 17 (6), 441-458.
- Konu, H., Laukkanen, T., & Komppula, R. (2011). Using ski destination choice criteria to segment Finnish ski resort customers. *Tourism Management*, 32(5), 1096-1105.
- Kotler, P., & Keller, K. (2006). *Marketing Management* (12th ed.). Upper Saddle River, New Jersey: Pearson Prentice Hall.
- Mackellar, J. (2009). Dabblers, fans and fanatics: Exploring behavioural segmentation at a special-interest event. *Journal of Vacation Marketing*. 15(1), 5-24.
- Mazanec, J. (2000). *Market Segmentation*. In: Encyclopedia of Tourism. J. Jafari, (Ed.) London: Routledge.
- Middleton, V., Fyall, A., Morgan, M., & Ranchhod, A. (2009). *Marketing in travel and tourism* (4th ed.). Slovenia: Butterworth-Heinemann.
- Molera, L., & Albaladejo, I. (2007). Profiling segments of tourists in rural areas of South-Eastern Spain. *Tourism Management*, 28(3), 757-767.
- Morrison, A. (2002). *Hospitality and Travel Marketing* (3rd ed.). USA: Delmar Publications.
- Moscardo, G., Pearce, P., & Morrison, A. (2001). Evaluating Different Bases for Market Segmentation: A Comparison of Geographic Origin versus Activity Participation for Generating Tourist Market Segments. *Journal of Travel & Tourism Marketing*, 10(1), 29-49.
- Oh, J., & Schuett, M. (2010). Exploring Expenditure-based Segmentation for Rural Tourism: Overnight Stay Visitors Versus Excursionists to Fee-fishing Sites. *Journal of Travel & Tourism Marketing*, 27(1), 31-50.
- 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. (2012). Segmentation of rural tourists: combining push and pull motivations. *Tourism and Hospitality Management*, 18(1), 69-82.
- Pesonen, J., & Komppula, R. (2010). Rural Wellbeing Tourism: Motivations and Expectations. *Journal of Hospitality and Tourism Management*, 17(1), 150-157.
- Sally, D., & Simkin, L. (2010) Judging the quality of customer segments: segmentation effectiveness. *Journal of Strategic Marketing*, 18(2), 113-131
- Saunders, J. (1994). Cluster Analysis. *Journal of Marketing Management*, 10(1), 13-28.
- Wedel, M., & Kamakura, W. (1998). *Market Segmentation-Conceptual and Methodological Foundations*. Boston: Kluwer Academic Publishers.
- Wilkie, W.L. (1994). *Consumer Behavior*. (3rd ed.). New York: John Wiley and Sons.