

## Market segmentation in tourism destination sector using self-organizing maps and its validation with structural equation modeling

### Segmentação de mercado no setor de destino turístico usando mapas auto-organizados e sua validação com modelagem de equações estruturais

DOI: 10.34140/bjbv4n2-004

Recebimento dos originais: 20/01/2022

Aceitação para publicação: 31/03/2022

#### Teófilo Câmara Mattozo

Doutor em Engenharia Elétrica e Computação pela Universidade Federal do Rio Grande do Norte  
Endereço: Rua Dr. João Chaves 940, CEP 59020-310. Apt. 602, Bairro Tirol, Natal- RN,  
Instituição: Instituição Universidade Estadual do Rio Grande do Norte  
E-mail: mattozo@oi.com.br

#### Gutemberg Soares da Silva

Doutor em Engenharia Elétrica e Computação pela Universidade Federal do Rio Grande do Norte  
Instituição: Universidade Federal do Rio Grande do Norte  
Endereço: Rua Miguel Barra 800, CEP 59.014-590. Apt. 1200, Bairro Tirol, Natal- RN, Brasil  
E-mail: guttembbergue@gmail.com.br

#### André Pedro Fernandes Neto

Doutor em Ciências da Computação e Engenharia Elétrica pela Universidade Federal do Rio Grande do Norte  
Instituição: Universidade Federal Rural do Semi-árido.  
Endereço: Rua João da Escócia 1727, CEP 59.607-330. Alphaville Lote K1 22 Mossoró-RN, Brasil  
E-mail: andrepedro@ufersa.edu.br

#### José Alfredo Ferreira Costa

Doutor em Engenharia Elétrica e Computação pela Universidade Estadual de Campinas  
Instituição: Universidade Federal do Rio Grande do Norte  
Endereço: Rua Nival Câmara, CEP 59.020-630. Apt. 602, Bairro Tirol, Natal- RN, Brasil  
E-mail: jafcosta@gmail.com

#### ABSTRACT

Nowadays, the tourism industry is growing up and competing relatively high, so how to make marketing strategies has been an essential issue for your market. A common problem for marketing strategists is how to appropriately segment the market and select segment-specific marketing strategies. Current segmentation methodologies and clustering techniques have evolved to include the use of artificial neural networks to the segmentation of markets. This article presents an approach that combines the Self-Organizing Map of Kohonen (SOM) with the technique Structural Equation Modeling (SEM) in the application of a segmentation issue. A satisfaction model is used with SEM to measure customers' habits and attitudes, in Brazilian tourist destination also looking at the effects of moderating segments. These market maps not only help the managers to see fully visualized clusters of market but also reveal mutual correlations between different customers' characteristic variables. SOM algorithm and visualizing technique were implemented in MATLAB environment to produce market maps of data set.

**Keywords:** Market Segmentation, Tourist Destination, Self-Organizing Maps (SOM), Structural Equation Modeling, Moderating Effect of the Segmentation

## RESUMO

Atualmente, a indústria do turismo está crescendo e competindo em alto nível, assim, como fazer estratégias de marketing tem sido uma questão essencial para seu mercado. As atuais metodologias de segmentação e técnicas de agrupamento evoluíram para incluir o uso de redes neurais artificiais para segmentação de mercado. Este artigo apresenta uma abordagem que combina os mapas auto-organizáveis (SOM) de Kohonen com a técnica de modelagem de equações estruturais (SEM). Este artigo apresenta uma abordagem que combina o Mapa Auto-Organizável de Kohonen (SOM) com a técnica de Modelagem de Equações Estruturais (SEM) na aplicação de um problema de segmentação. Um modelo de satisfação é usado com SEM para medir os hábitos e atitudes dos clientes, no destino turístico brasileiro observando também os efeitos da moderação dos segmentos. Esses mapas de mercado não apenas ajudam os gerentes a distinguir clusters de mercado totalmente visualizados, mas também revelam correlações mútuas entre as variáveis características de diferentes clientes. O algoritmo SOM e a técnica de visualização foram implementados em ambiente MATLAB para produzir mapas de mercado do conjunto de dados.

**Palavras-chave:** Segmentação de Mercado, Destino Turístico, Mapas Auto-Organizáveis (SOM), Modelagem de Equações Estruturais, Efeito Moderador da Segmentação.

## 1 INTRODUCTION

Tourism is one of the most important industries for economic development and job. It can make revenue is causing a turnover in the country. To be used in tourism planning and managing to design products or services in accordance with the needs of tourists. Therefore, tourism organizations need to understand the tourism patterns and preferences of tourists. In a globalized world, where to be differentiated becomes more important every day the tourists require, increasingly, tours that fit their needs, their personal situation, their desires and preferences. Thus, the segmentation has a significant role as a strategy for the structuring and the marketing of the tourism destinations and itineraries. Thus, for the segmentation of the tourism to be effective, it is necessary to deeply understand the characteristics of the target: the offer (attractions, infrastructure, tourism products and services) and demand (the specificities of the groups of tourists who have already visit or will visit).

The value of performing marketing segmentation analysis includes better understanding of the market to properly position a product in the marketplace, identifying the appropriate segments for target marketing, finding opportunities in existing markets, and gaining competitive advantage through product differentiation [1].

In Brazil there is little information about tourism market segmentation, existing unexplored dimensions related to the factors that influence the segmentation and practically nothing about its validation with structural and measurement models [2]. Although a number of clustering methods have been presented to resolve the issue of market segmentation, the importance of testing their validity is often ignored by the marketing researchers [3]. If the identified segments are inadequately validated, the results can be considered devoid of reliability.

The main contribution of this research is the combination the SOM algorithm with the SEM

technique in the application of market segmentation issue. The SOM is used to view the market segments and SEM is applied on the validity test of equivalence of a same theoretical structure. The verification of the structural model was performed by means of the steps involved in multiple group invariance tests, investigating the various features of the tourists.

The objectives of this article are: (1) use self-organizing maps (SOM) in the market segmentation of tourist destination process, based on demographics data, attitudes and behaviors of the customers; (2) validate the segmentation with structural models and (3) examine how the moderation effects of the segmentation performed can affect the assessment of overall satisfaction, especially the relationship with their background, quality, value and image. The next section presents the theoretical foundations and the model used. It presents a brief introduction of the SEM technique, as well as, a description of the methodological procedures performed in this study. Next, the customers are segmented using SOM [4, 5], followed by the validation tests with multiple analysis of the measure and means models with the proposed model. Finally, the moderating effects of the customer segmentation in the structural model. After the presentation of the results, a discussion and recommendations are presented.

## 2 CONCEPTUAL MODEL AND STRUCTURAL EQUATIONS

### 2.1 CONCEPTUAL MODEL

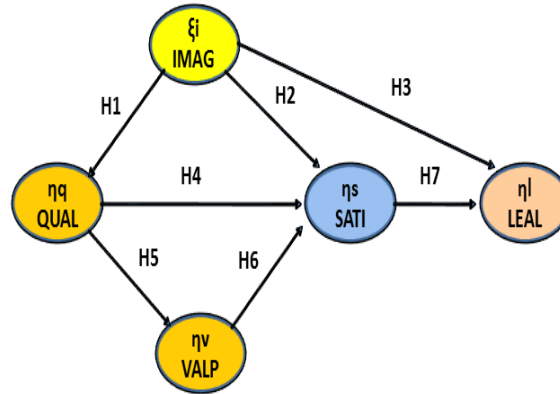
For the present study, it was adopted the methodology of those authors, who considered various constructs as antecedents and consequences of satisfaction. The theoretical and empirical methods used in this study are mainly based on those of Fornell *et al.* [6]. Thus, there were considered five constructs correlated to the tourist destination, in order to operationalize the proposed conceptual model:

- 1 - Loyalty: includes indicators of more emotional and behavioral nature, which influences the evaluation of the visitor about the likelihood of revisiting the same destination or disposition to recommend it to others.
- 2 - Overall satisfaction: central variable in the model, resulting from the travel experience of visitors to satisfy their desires, expectations and needs in relation to the trip.
- 3 - Perceived value: the relation between quality and price, represents the evaluation of the visitors of what is received (benefits), and what is given (costs or amount paid).
- 4 - Quality of the Travel: defined as the visitor judgment about the superiority or excellency of the services related to the travel experience.
- 5 - Destination image: integrates all kind of associations that the visitor makes about the tourist destination.

The explanatory theoretical model of the relationship of the latent variable satisfaction with its antecedent and consequent variables was developed based on the literature review with the foundations and related assumptions presented in Mattozo [7]. The graphical representation of the model of satisfaction

of the customer tourist destination involving the latent variables that influence or are influenced by satisfaction, is presented in Fig. 1.

Figure 1. Structural Diagram of the Conceptual Model



Source: Mattozo [7].

## 2.2 STRUCTURAL EQUATIONS OF THE MODEL OF CUSTOMER SATISFACTION OF TOURIST DESTINATION

The models based on structural equation can be understood as models of research based on statistics, correlations and regression analysis, including independent and dependent latent variables and measurement errors in the variables, multiple parameters, reciprocal causes, simultaneity and interdependence [8]. For Bollen [9], a SEM with latent variables is set according to the equations:

$$\eta = \alpha + B\eta + \Gamma\xi + \zeta \quad (1)$$

$$y = \mu_x + \lambda_y\xi + \varepsilon \quad (2)$$

$$x = \mu_y + \lambda_x\xi + \delta \quad (3)$$

In the structural model proposed in Fig. 1, the image represents an exogenous variable ( $\xi_i$ ), the endogenous variables are quality ( $\eta_q$ ), value ( $\eta_v$ ), satisfaction ( $\eta_s$ ) and loyalty ( $\eta_l$ ). The elements of the matrix B represent the direct causal effects of  $\eta$  in other variables  $\eta$  and the  $\gamma$  elements of the matrix represent the direct effect of the variable image  $\xi_i$  in the endogenous variables  $\eta$ . The  $\Phi(n \times n)$  and  $\Psi(m \times m)$  matrices, not represented in the equation, are the matrices  $\xi$  and  $\zeta$  respectively.

$$\begin{bmatrix} \eta_q \\ \eta_v \\ \eta_s \\ \eta_l \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \beta_{vq} & 0 & 0 & 0 \\ \beta_{sq} & \beta_{sv} & 0 & 0 \\ 0 & 0 & 0 & \beta_{ls} \end{bmatrix} * \begin{bmatrix} \eta_q \\ \eta_v \\ \eta_s \\ \eta_l \end{bmatrix} + \begin{bmatrix} \gamma_{qi} \\ 0 \\ \gamma_{si} \\ \gamma_{li} \end{bmatrix} * \xi_i + \begin{bmatrix} \zeta_q \\ \zeta_v \\ \zeta_s \\ \zeta_l \end{bmatrix} \quad (4)$$

The adjustment methods used in the analysis of structural equations and the functions of discrepancies vary according to the software used and the assumptions about the nature of the measures, as well as the distribution featuring the variables. The function used in this study and, according to Hair *et al.* [10], the most frequently used is the maximum likelihood (ML), which seeks to estimate the model by means of a maximum likelihood function with the theoretical parameters, for which the sample collected identifies the maximum probability of occurrence [11], and is defined by:

$$F_{ML}(S, \Sigma(\theta)) = \log |\Sigma(\theta)| + \text{tr}(S \cdot \Sigma(\theta)^{-1}) - \log |S| - (p + q) \quad (5)$$

where  $p$  and  $q$  are respectively, the quantity of endogenous and exogenous variables manifested. For estimation of means of the variables, the function must include the term:

$$[\bar{x} - \mu(\hat{\theta})]' \Sigma(\hat{\theta})^{-1} [\bar{x} - \mu(\hat{\theta})] \quad (6)$$

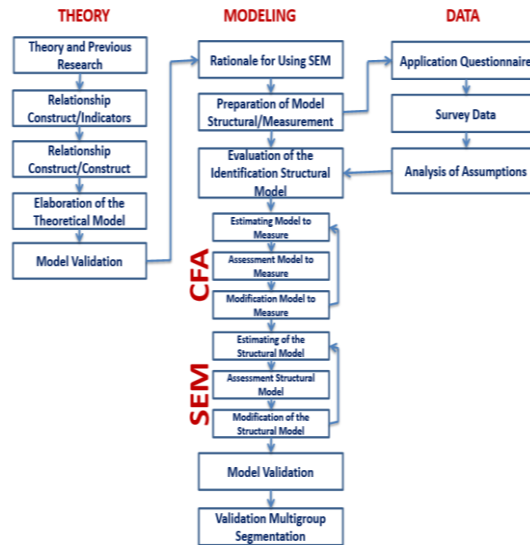
where  $\bar{x}$  represents the mean of the variables manifested and  $\mu(\hat{\theta})$  represents the vector of mean estimated by the model.

### 3 METHODS

The development of an integrated tourist destination system is a complex task, especially for the non-deterministic characteristic of this type of system. Therefore, it is indispensable to use a complete and systematic methodology. A methodology is often presented as a series of steps, with techniques and notation associated with each step [12].

The methodology proposed for the development of this research is recurring to the exploration of the concepts and characteristics advocated by Bollen [9], Kline [8], Schumacker and Lomax [13], Hair *et al.* [10] and Marôco [14]. The development of these components, integrated into a tourist perspective, allows the definition of a strategy for the SEM application, comprising in three phases: Theoretical Study, Structural Modeling and Data Analysis (Fig. 2).

Figure 2. Research methodology steps.



Source: Mattozo [7].

### 3.1 DETERMINATION OF THE POPULATION, SAMPLING, SURVEY, INSTRUMENT AND TECHNIQUE OF DATA COLLECTION

The research was conducted in the city of Natal, capital of Rio Grande do Norte/Brazil. The study population was composed of tourists who used the services offered by this tourist destination. With the population matrix identified, it was used the systematic random sampling technique, in a field survey conducted between January and March of 2013 with a questionnaire application, by a staff trained by the researcher. The process used to obtain the samples needed for the satisfaction research was structured through a sampling plan with a confidence interval of 95% and a maximum error of 5% [15]. The sample size calculation was performed based on the estimation of proportions. To achieve the confidence level and the margin of tolerable error chosen, the sample was calculated using the equation described in Larsson and Farber [16], 456 questionnaires were obtained, of which 420 were considered valid (difference established due to the assumptions of normality, linearity and multicollinearity of the SEM).

It was defined a panel of fifteen experts consultants, composed of professors of the Tourism course of the State University of Rio Grande do Norte and managers of recognized reputation in the major hotels and tourist agencies of Natal. There were conducted personal interviews using a questionnaire with two types of variables: related to the content and of characterization. A list of 33 measuring items was developed using the input information, formulated from the review related to the study presented by Mattozo [7] literature. The constructs of the model were measured using a questionnaire of multiple scales items. The measures used a response format of 10 points Likert.

### 3.2 PROCEDURES OF MARKET SEGMENTATION WITH SOM

Methodologies of market segmentation and clustering techniques have evolved to include the use of artificial neural networks. The self-organizing maps (SOM) can project the input space into a high dimensional in a low-dimensional topology, allowing it to be possible to visually determine the clustering [17]. The basic SOM training algorithm consists of three phases [5]. In the first phase, competitive, there are selected the parameters of the map such as the dimensions and weights of the initialization vector, corresponding to each neuron. In the second phase, the cooperative, it is defined the neighborhood of this neuron. The network is fed with data in analysis to find the best matching unit for each vector of input data. Each X record includes quantitative values of n attributes that are shown as:

$$X = [X_1, X_2, \dots, X_n] \in R^n \quad (7)$$

the weight vector of the i neuron is defined as:

$$m_i = [m_{i1}, m_{i2}, \dots, m_{in}] \in R^n \quad (8)$$

Then, in correspondence to each input register, the best associated unit, ie, the winner neuron is identified based on the equation:

$$c = \arg \min_i \{d(X, m_i)\} \quad (9)$$

where c indicates the winner neuron and  $d(X, m_i)$  is the Euclidean distance between the record and the weight vector of the ith neuron, which is calculated by the equation:

$$D(X, Y) = \|X - Y\| \quad (10)$$

In the last phase, called adaptive, the vector weight is updated corresponding to each neuron, using the equation:

$$m_i(t+1) = m_i(t) + \alpha(t) h_{ci}(t) [X(t) - m_i(t)] \quad (11)$$



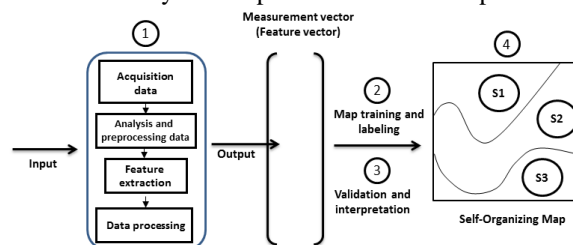
where,  $0 < \alpha < 1$  is the learning rate and  $h_{ci}(t)$  indicates the rate of the neighborhood of the  $i$ th neuron with neuron of order  $c$  (winner neuron). The rate of neighborhood of the  $i$ th neuron with the winner neuron is obtained from the equation below, which is a Gaussian function.

$$h_{ci} = e^{-\frac{\|r_c - r_i\|^2}{2\sigma^2(t)}} \quad (12)$$

where  $\sigma$  is the domain controller function that is, eventually, decreasing during the training process, with  $r_i$  and  $r_c$ , respectively, the positions of the neurons of the  $i$  and  $c$  order, in the self-organizing map SOM [5].

A schematic illustration of the application of the SOM analysis, in a process of service experience, adapted from Simula *et al.* [18], is shown in Fig. 3. The different stages of the figure are: (1) the data processing (acquisition, preprocessing, feature extraction and normalization), (2) training map, (3) validation and interpretation and (4) data visualization.

Figure 3. Steps of the SOM application in the analysis of a process of service experience.



Source: Adapted from Simula *et al.* [18].

After conducting empirical tests, among the various settings of the SOM used, it was chosen the flat topology with hexagonal grille, Gaussian neighborhood function and training algorithm in batch. This topology was chosen due to lower processing time and grid size of 40 x 40 neurons. The algorithm was set to last for 10,000 times and has an end radius of zero growth. In the second step, the K-means algorithm is used for various values of  $k$  by using the map of the matrix [19]. Each result is evaluated by the Davies-Bouldin index, which is a function of the ratio of the sum of the intergroup distances for the distance between groups and in which the best segmentation is the one with the lowest index [20]. After determination of the clusters, the output nodes to the input data, called BMU's (Best Matching Unit) are used to classify the incoming data, ending the segmentation.

Following the guidelines referenced by Kuo *et al.* [21], there were used to evaluate the segmentation model: 70 % of the population for training and 30% for testing. In the 420 observations of the potential market, a sample of 294 was used for training and the remaining 126 observations, for testing.



### 3.3 TEST PROCEDURES FOR VALIDATION OF THE MARKET SEGMENTATION

The tests to the validity of the hypothetical model and measurement invariance, in all groups, are necessary conditions for a meaningful and precise comparison between groups in the constructs of interest [22]. With these two conditions met, the test for significance of differences between groups on the constructs can be efficiently performed within SEM [9]. The validity of the conceptual model should be tested separately for each group of customers.

A Confirmatory Factor Analysis (CFA) under SEM was performed using AMOS<sup>®</sup> V.19 software for the statistical analysis. To assess if the factorial model proposed is invariant under test between the groups, the parameters sets related to the factor weights and the covariance between the factors, are evaluated in a ordered and sequentially restrictive manner, carrying out a comparison of the adjustment of different groups simultaneously, considering: the factor weights and the covariance of the measurement model; the structural coefficients of the causal relationships; the variances and covariances of the errors.

The function of variance for the multiple group analysis is given by:

$$F = \sum_{m=1}^M \binom{n_m}{N} f_m (S_m \sum_m W_m) \quad (13)$$

where  $f_m$  is the discrepancy function defined for the analysis [11]. The analysis of invariance seeks to demonstrate that the factor model for each of the groups, the factor weights and covariances, do not differ significantly, following the formulation below:

$$x = \tau_x + \Lambda_x \xi + \delta \quad (14)$$

where  $\tau_x$  represent a vector of intercepts of the variables  $x$ ,  $\Lambda_x$  refers to the matrix of load factors, the vector  $\xi$  of the latent variable scores and  $\delta$  is the vector of the error terms from the corresponding measurements of the variables  $x$ . The difference in chi-square ( $\chi^2$ ) is the statistic test for assessing the factorial invariance of all the groups, between the fixed and free models, with degrees of freedoms previously defined.

$$\chi_{Dif}^2 = \chi_{Invar}^2 - \chi_{NInvar}^2 \quad (15)$$

To perform the test of invariance of the factor weights of the measurement model, equality restrictions are applied to the weights of all groups. As a condition for testing factor invariance, the reference model is considered, which is estimated for each group separately, selecting the one that best fits the data, based on the parsimony and relevant significance views. The most parsimonious,

substantially more significant model is referred to as the model for this group [14]. With the assumption of invariance of the measurement model met, the path for comparing the latent variables is established. Similarly, the analysis of invariance applies to the coefficients of the structural model. The analysis of multiple groups can also be performed by comparing the structural paths between latent variables in a full structural equation model. The analysis of the means between groups can also be used to test significant structural differences between the groups, with models that contain intercepts [22].

## 4 DATA ANALYSIS AND RESULTS

### 4.1 ANALYSIS OF THE MARKET SEGMENTATION

The SOM algorithm and visualization techniques were implemented in MATLAB® V.7 environment to map a set of market data, the result of the segmentation consists of training and testing samples. Through the application of the segmentation technique used, the results showed the existence of three well-defined clustering, based on various demographic characteristics, socioeconomic factors and the importance that the respondents assigned to each behavioral factor when assessing the latent variables. This information can provide with a clear view of the market segments and their specifications for market managers who are generally interested in recognizing their target segment or segments.

The first segment (S<sub>1</sub>) included 158 members, the second (S<sub>2</sub>) 85 members and third (S<sub>3</sub>) 177 members. After the identification of the number of resulting clustering, it was proceeded its characterization based on the revealed preferences. Indeed, there were identified (table 1) the average scores obtained in the importance set by the tourists, in the three clustering obtained, by the SOM algorithm.

Aiming to characterize each segment, there was observed the means obtained in the importance placed by the customers, considering each of the factors. The customers of each segment have particular characteristics.

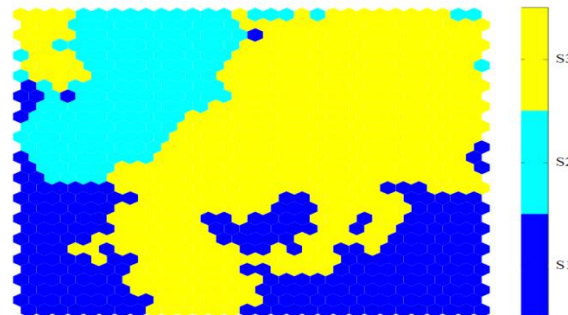
Table 1. Characterization of the segments

Variables		Segments		
		S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>
Characterization	Gender	Men	Men	Women
	Travel with	family	family	family
	Number of visits	Once	Once	Once
	Purpose of Travel	Leisure	Leisure	Leisure
	Occupancy	Private companies	Entrepreneur	Private companies
	Permanence time	4 to 7 days	4 to 7 days	1 to 3 days
	Age	Adults	Seniors	Young
	Schooling	High education	High education	High education
	Social class	B	A	B
	Content	Image	7.3	8.2
Quality		7.6	8.3	5.7
Value		5.7	8.7	4.6
Satisfaction		7.0	8.5	4.8
Loyalty		7.1	8.4	4.7

Source: Authors with MATLAB®

The results presented in Fig. 4 and 5 illustrate that the clusterings have a correct variability and are well-differentiated. The Fig. 4 shows the tourist market searched, where each set of the map corresponds to a segment, considering three different segments may be identified. The U matrix shows, in different colors, the segments that make up the market.

Figure 4. Segmented matrix U-Clusters are identified by colors



Source: Authors with MATLAB®

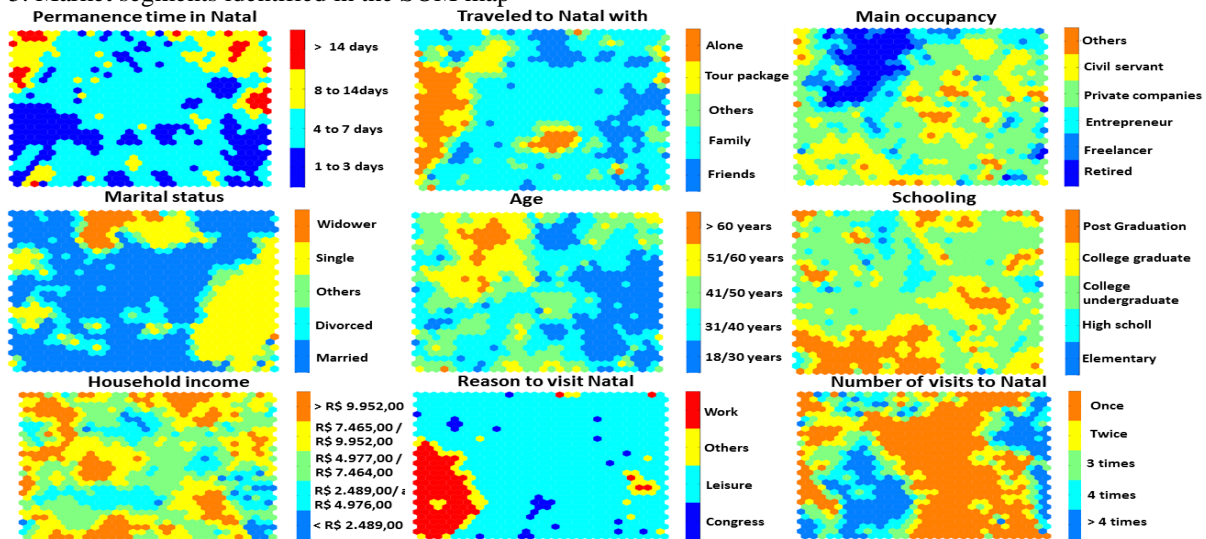
The members of the  $S_1$  segment (38% of the sample) are predominantly adults, male and middle high class, with greater permanence among the groups, have a good image of the destination and cares about the price. The components of the  $S_2$  segment (20 % of the sample) have the following outstanding characteristics: senior, married, entrepreneurs, belonging to the upper class, and leisure travelers in the company of the family. They have a great image of the destination and expressed great satisfaction with the trip. The  $S_3$  segment (42 % of the sample), which mostly, consists of young, single women, visiting the destination for the first time, with the shorter permanence among the groups. They expressed dissatisfaction with the quality of services, are very price sensitive and have low loyalty.

In Fig. 5, each variable map shows the distribution of the values of the corresponding variable, revealing the underlying knowledge of the tourists' data.

Some considerations that may be inferred from the variable maps are:

1. Schooling and income are positively correlated in the market previous mentioned. In other words, anywhere in the map, where there is a high level of education (brown area in the schooling map), there is a correlation with the high level of income (brown area in the income map), and vice versa.
2. Most of the retirees tourists (dark blue of the occupancy map), used the option of tour packages (yellow area of the destination map) and had a high residence time (light blue area in the permanence map).
3. Tourists characterized as civil servants (light green area of the occupancy map) provided single (yellow area in the marital status map) had a low permanence time (dark blue area of the map).

Figure 5. Market segments identified in the SOM map



Source: Authors with MATLAB®

#### 4.2 TESTS FOR INVARIANCE OF THE MEASUREMENT MODEL AMONG THE SEGMENTED GROUPS

The AFC was performed with AMOS® V.19 statistical software [23]. The verification of the assumptions required to perform the analysis of the structural equations, the coefficients of consistency and the fit indices of the factor model, demonstrating the validity of the model used to assess satisfaction, as well as the adjustment of the parameters of the overall model, were developed in Mattozo [7]. The validity of the factorial model was tested by AFC for the groups presented together, in the model shown in Fig. 4. The following indices of adjustment were calculated: Chi-square ( $\chi^2$ ), the Comparative Fit Index (CFI), the Goodness of Fit index (GFI), the index of Goodness of Adjusted Fit (AGFI), the Index of Adjustment Not Normalized (NNFI), the Mean Squared Error Approximate (RMSEA), with a confidence interval of 90% and the Information Criterion Akaike (AIC). A great adjustment of Goodness was obtained with CFI=0,931, GFI=0,910, AGFI=0,903, NNFI=0,926, RMSEA=0.035 and AIC=790,775, simultaneously to the entire sample.

The restrictions of parameters between segments were created in consonance with the recommendation of Byrne [22], generating four models: Model M<sub>0</sub> - No restriction (all parameters free); M<sub>1</sub> - Factorial weights (equal factor weights); Model M<sub>2</sub> - Structural Covariance (equal weights and covariances between factors) and Model M<sub>3</sub> - Measurement of Waste (weights, covariance and waste, all the same).

The invariance of the measurement model was initially performed to compare three groups of S<sub>1</sub>, S<sub>2</sub> and S<sub>3</sub> segments, composing the entire sample. The comparison of the free model with the fixed factor weights is given by the first line of table 2, which shows the output data of the statistic test generated by AMOS® V.19 software. It can be seen that the degree of freedom (df) with 52 yielded a value  $\chi^2 = 78,392$ ; p=0,058. Referring to  $\chi^2$  distribution table, it appears, to a level of statistical significance of 0.05, a relation

between the theoretical value with the measured of  $\chi^2_{0,95} (52) = 69,832 < 78,392$ . Thus, the hypothesis that the  $M_1$  can be as well-adjusted as  $M_0$  is rejected. Therefore, it is demonstrated the variance of the factor weights in three segments.

Following, in sequence, the same procedure of rating the difference of the  $\chi^2$  for the evaluation test of the invariance factor between the fixed and free models, with the degrees of freedom defined, it was compared the adjustment of  $M_2$  e  $M_3$ , respectively, with  $M_0$ , among the three segments. Considering the same statistical significance of 0.05 and rating of the difference of  $\chi^2$ , it can be concluded that the quality of adjustment of the two models are also significantly different in view of the statistical relationships obtained, confirming the lack of equivalence between the three segments tested.

Finally, using the same hierarchical analysis strategy, of the simpler models (all free parameters) for the more complex models (all fixed parameters), there were, sequentially, performed comparisons of customer segments combined two by two:  $S_1$  and  $S_2$ ,  $S_1$  and  $S_3$ ,  $S_2$  and  $S_3$ , utilizing the difference test  $\chi^2$ , as reported in table 2.

The results confirm a significant difference between all segments analyzed, fully validating the segmentation performed.

Table 2. Statistics of the difference of  $\chi^2$  for the free and fixed models

Segments compared	Models	df	Measured Value		Tabular Value		Invariance
			$\chi^2$	P-Value	$\chi^2$	P-Value	
$S_1/S_2/S_3$	$M_1$	52	78,392	0,058	69,832	0,05	NOK
	$M_2$	90	176,762	0,155	113,145	0,05	NOK
	$M_3$	173	357,437	0,079	204,611	0,05	NOK
$S_1/S_2$	$M_1$	28	49,756	0,019	41,337	0,05	NOK
	$M_2$	48	65,705	0,016	65,171	0,05	NOK
	$M_3$	87	130,605	0,009	109,773	0,05	NOK
$S_1/S_3$	$M_1$	28	48,466	0,028	41,337	0,05	NOK
	$M_2$	48	64,186	0,042	65,171	0,05	NOK
	$M_3$	87	147,219	0,054	109,773	0,05	NOK
$S_2/S_3$	$M_1$	28	47,253	0,037	41,337	0,05	NOK
	$M_2$	48	63,657	0,025	65,171	0,05	NOK
	$M_3$	87	234,054	0,089	109,773	0,05	NOK

Source: Prepared by the authors with AMOS®

### 4.3 COMPARISON OF MEANS OF THE STRUCTURAL MODEL CONSTRUCTS AMONG THE SEGMENTED GROUPS

The structural analysis of means between groups is applied to models containing the ordinate (intercept) in the origin for the purpose of estimating the mean for a specific group construct. In the structural analysis of the model, the mean of each observed variable  $X_k$  is obtained from means of the latent variables in each group, according to equation [24]:

$$X_k = \tau_k + \lambda_k [\text{mean}(\xi)] \quad (16)$$

For this analysis, two models were generated and evaluated sequentially, among  $S_1$ ,  $S_2$  and  $S_3$  segments combined two by two. In the first model ( $M_L$ ), the factor weights were set to zero and the second ( $M_F$ ), in addition to the weights, the intercepts were also fixed. The means of the residues of the segments were fixed at zero. In the first evaluation, considering the  $S_1$  and  $S_2$  segments, the average of the  $S_1$  segment was also set to zero, making it the reference group. To make the difference of the  $\chi^2$  it was proceeded to adjust the model ( $M_F$ ), in both segments in the same value. After the adjustment of the model, there are shown in table 3, the  $\chi^2$  values for  $M_L$  and  $M_F$  models of three segments, two by two. The difference of the two models was  $\chi^2_{Diff} = 1.254,178 - 1.205,447 = 51,731$  with  $df (872-865)=7$  degrees of freedom, and observing the statistical relationship  $\chi^2_{095}(7)=14,067 < 51,731$ . It appears therefore that the  $S_1$  and  $S_2$  segments are significantly different confirming the lack of equivalence among the tested segments and consequently the validity of the proposed segmentation.

Similarly, extending the application of the test of the difference of  $\chi^2$  to  $M_L$  and  $M_F$  models, for the other combinations of segments procedure, it appears that there are also significant differences among the means of the segments in the constructs of the structural model. Therefore, these results ratify the confirmation of validation of the segmentation performed by the Kohonen map, presented in the previous section.

Table 3. Statistics of the difference of  $\chi^2$  for the free and fixed models

Segments compared	Models	df	Measured Value		Tabular Value		Invariance
			$\chi^2$	P-Value	$\chi^2$	P-Value	
$S_1/S_2$	$M_1$	7	51,731	0,161	14,067	0,05	NOK
$S_1/S_3$	$M_1$	7	25,557	0,074	14,067	0,05	NOK
$S_2/S_3$	$M_1$	7	26,311	0,068	14,067	0,05	NOK

Source: Prepared by the authors with AMOS®

#### 4.4 ANALYSIS OF THE EFFECT OF MODERATION SEGMENTS IN THE STRUCTURAL MODEL

Baron and Kenny [25], define a moderator variable as a quantitative or qualitative variable, that affects the direction and/or strength of the relationship between the independent and dependent variables. In order to test the moderating effect of the targeted customer groups, the proposed structural model was used, in the analysis of the multi group structural equation. For this purpose, three groups were generated with the segmentation procedure described in item 3.2, motivating the creation of a new variable in the database, resulting in the identification of the  $S_1$ ,  $S_2$  and  $S_3$  segments. The three relationships with the satisfaction (image, quality and value) while moderating variable segmentation tourists were tested. The results are shown in table 4.

In the first relationship, Quality  $\rightarrow$  Satisfaction, the effect occurs in the same direction of the satisfaction, the coefficient of the highest value (0.36) was identified for the group  $S_1$ , compared to the groups  $S_2$  (0.08) and  $S_3$  (0.09). In other words, the impact of quality on satisfaction was significantly



higher in the group of tourists, among other aspects, were characterized, by expressing a strongly positive assessment on the quality of service, who, mostly, are men, work in private companies, adults, social class B, in addition to a high education, as shown in table 1 of section 4.1.

Table 4. Ratios between the antecedents of moderate satisfaction with the segments

Relationship between Constructs	Segments	$\beta$ Standardized	$\beta$ Not standardized	p-value
Quality → Satisfaction	S <sub>1</sub>	0,36	0,470	0,032
	S <sub>2</sub>	0,08	0,09	0,841
	S <sub>3</sub>	0,09	0,07	0,884
Image → Satisfaction	S <sub>1</sub>	0,23	0,29	0,017
	S <sub>2</sub>	0,12	0,14	0,438
	S <sub>3</sub>	0,41	0,49	0,016
Value → Satisfaction	S <sub>1</sub>	0,35	0,28	0,009
	S <sub>2</sub>	0,54	0,46	0,032
	S <sub>3</sub>	0,38	0,29	0,006

Source: Prepared by the authors with AMOS®

In the second relationship, Image → Satisfaction there was greater coefficient in S<sub>3</sub> (0.41) group compared to the groups S<sub>1</sub> (0.23) and S<sub>2</sub> (0.12). This reveals that the effect of the image on satisfaction is stronger for tourists of the S<sub>3</sub> segment. This result is consistent with what was found in the evaluation of the profile associated with this segment of tourists, ie, values mostly above average in the judgment of the respondents to this construct, as shown in section 4.1. They are, mostly, females, working in the private sector, are young, from the social class B with high schooling and visited this destination for the first time with length of stay of 1-3 days. They are also the most demanding.

In the third relationship, Value → Satisfaction, it was identified as the highest value of coefficient S<sub>2</sub> (0.54), although with slightly superior to the segment S<sub>1</sub> (0.35) and S<sub>3</sub> (0.38). It appears that the effect of the value on the satisfaction was found with greater relevance in the segment whose tourists involved expressed a high evaluation to this construct, which together with their differential characteristics, gave rise to the identification of the group, because they have high income, are male, aged over 50 years, are entrepreneurs and traveled with the family, mostly. They are also the most tolerant.

These results are in line with several previous studies where it was evident that the characteristics of the customers have moderating effects on the results of their satisfaction, repurchase intention as and communication mouth-to-mouth [26, 27]. Also corroborate with these results, Bryant and Jaesung [28] and Mittal *et al.* [26], that the characteristics of customers, such as gender, age and educational level, among others, have a major impact on the level of customer satisfaction. Assaker and Hallak [29], investigated the moderating effects of the trends of the tourists in search of news, in the relationship between destination image, satisfaction and intentions to revisit in short and long term. Nicolau and Mas [30], found that motivation moderated the relationship between the service quality, satisfaction and loyalty in the tourist destination. Perillo [31], observed differences associated with some components of a model of customer satisfaction between men and women, in banking services. Linn [32], examined differences



in the satisfaction of burned patients, identifying moderating effects between groups of patients, considering the age and the severity of the case.

## 5 CONCLUSIONS

Marketing segmentation plays a relevant role in modern marketing paradigms, but little has been done to make it an effective instrument for visualizing market segments in a way which not only easily understood by all market managers but can also show the knowledge underlying the available dataset.

In this study, an integrated approach of SOM algorithm with the SEM technique in application to the problem of market segmentation was developed. Content variables defined by a model customer satisfaction, in the tourism area, besides the characterization variables, were obtained by a survey forming the data set for training the network. After building and training the SOM, the U matrix and the maps of the variables were produced, revealing not only the existence of three market segments, but also the mutual relations among all variables.

The members of the  $S_1$  segment are predominantly adult, male, high middle class with the longest length of stay between the groups, have a good image of the destination and care about the price. The components of the  $S_2$  segment have the following outstanding characteristics: senior, married, entrepreneurs belonging to the upper class, and leisure travelers in the company of family, have a great destination image and expressed great satisfaction with the trip. The  $S_3$  segment, mostly, consists of young unmarried women, visiting the destination for the first time, with the shortest length of stay between the groups. They expressed dissatisfaction with the quality of the services, are very price sensitive and have low loyalty.

The results show that the SOM technique is a valid method for market segmentation and SEM has proved to be successful in validating segmentation with structural models, being investigated various demographic, socioeconomic and behavioral factors of tourists.

Furthermore, it was found that the moderating effect of market segments can affect the assessment of the overall satisfaction, especially the relationship with their background. The hypothesis that the relationship between the constructs quality and satisfaction, image and satisfaction, as well as value and satisfaction, are different between the tourists group was validated. In these relationships, the impact of quality on satisfaction was significantly higher in the group of tourists from the  $S_1$  segment. Moreover, the effect of the image in the satisfaction was stronger for tourists of the  $S_3$  segment. In the third relationship assessed, it was identified in the  $S_2$  segment, the impact of greater relevance of the construct value on the satisfaction.

Thus, a segmentation based on consumer preferences can allow the market researchers and professionals in the tourism sector, to know the market with greater accuracy and detail, and, consequently, develop marketing strategies more appropriate to the characteristics of each segment

obtained, increasing benefits. Despite the theoretical and methodological care, we must recognize that any empirical study of this nature always has limitations due, usually, either to methodological options adopted, or the manner in conducting the fieldwork. With regard to the quantitative techniques, which, although significantly stronger for extracting statistical information and their relationships, some interactions may not have been perceived, and thus not being able to capture all the differences in the human behavior, for more diverse and consistent that the methods are.

Another limitation concerns the sensitivity of the estimation techniques used in relation to the assumption of normality and the coefficients of kurtosis and skewness, whose results showed that it was a negative leptokurtic and asymmetric distribution, making it necessary to use transformations of the original data. Research efforts can be made, in order to extend the study longitudinally, collecting data in certain time intervals, in order to reduce possible effects of seasonality and consolidate the interrelatedness of the variables involved.

## REFERENCES

- [1] Drummond G., Ensor J., Ashford R. Strategic Marketing: Planning and Control. Third Edition. Ed. Elsevier. 2008
- [2] Mattozo, T. C. “Contribuições de modelagem de equações estruturais na análise de dados em modelos comportamentais de destino turístico”, Tese de doutorado, Brasil, Universidade Federal do Rio Grande do Norte, P. 150, 2014.
- [3] Pilling, B., Crosby, L. and Ellen, P. “Using benefit segmentation to influence environmental legislation: A bottle bill application”, *Journal of Public Policy and Marketing*, 10(2), pp. 28-46, 1991.
- [4] Kohonen, T., Oja, E., Simula, O., Visa, A. and Kangas, J. “Engineering applications of the self-organizing map”, *Proceedings of the IEEE*, 84(10), pp. 1358-1384, 1996.
- [5] Kohonen, T. “Self-organizing maps”, 3<sup>a</sup> ed., Springer Series in Information Sciences. Berlin, Germany: Springer-Verlag, pp. 37-190, 2001.
- [6] Fornell, C., Johnson, M. D., Anderson, E. W., Cha, J. and Bryant, B. E. “The American customer satisfaction index: nature, purpose, and findings”, *Journal of Marketing*, 60(4), pp. 7-19, 1996.
- [7] Mattozo, T. C. “Contribuições de modelagem de equações estruturais na análise de dados em modelos comportamentais de destino turístico”, Qualificação de doutorado, Brasil, Universidade Federal do Rio Grande do Norte, P. 150, 2012.
- [8] Kline, R. B., “Principles and practice of structural equation modeling”, 3<sup>a</sup> ed., New York: The Guilford Press, P. 427, 2010.
- [9] Bollen, K. A., “Structural Equations with Latent Variables”, 1<sup>a</sup> ed., New York: John Wiley & Sons, Inc, P. 514, 1989.
- [10] Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E. “Análise Multivariada de Dados”, 6<sup>a</sup> ed., São Paulo: Bookman, P. 688, 2009.
- [11] Jöreskog, K. G. and Sörbom, D. “LISREL 8: User’s Reference Guide”, Scientific Software International, *Journal of Marketing Research*, 28(1), pp. 142-157, 1996.
- [12] Mattozo, T. C. “Análise de desempenho de vendas em telecomunicações utilizando técnicas de mineração de dados”, Dissertação (Mestrado em Engenharia de Produção), Brasil, Universidade Federal do Rio Grande do Norte, P. 229, 2007.
- [13] Schumacker, R. E. and Lomax, R. G. “A beginner’s guide to structural equation Modeling”, 3<sup>a</sup> ed., New Jersey: LEA, P. 536, 2010.
- [14] Marôco, J. “Análise de Equações Estruturais: fundamentos teóricos, software & aplicações”, 1<sup>a</sup> ed., Report Number: Portugal, P. 384, 2010.
- [15] Malhotra, N. K. “Pesquisa de Marketing: uma orientação aplicada”. 4<sup>a</sup> ed., Porto Alegre: Bookman, P. 719, 2012.
- [16] Larsson, R. and Faber, B. “Estatística Aplicada”, 4<sup>a</sup> ed., Pearson Education do Brasil, São Paulo: São Paulo, P. 656, 2010.
- [17] Kuo, R. J., Ho, L. M. and Hu, C. M. “Cluster analysis in industrial segmentation through artificial

neural networks”. *Computers and Industrial Engineering*, 42, pp. 391-399, 2002.

[18] Simula, O., Alhoniemi, E., Höllmen, J. and Vesanto, J. “Monitoring and modeling of complex processes using hierarchical self-organizing maps”, In *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS'96)*, P. 737, 1999.

[19] Ultsch A. “Self-organizing neural networks for visualization and classification”, In: Opitz O., Lausen B., Klar R. *Information and Classification*, Springer-Verlag, Berlin, pp. 864–867, 1993.

[20] Davies D. and Bouldin D., “A Cluster Separation Measure”, *IEEE Trans. Pattern Anal. Mach. Intell.* 1(2), pp. 224-227, 1979.

[21] Kuo, R. J., An, Y. L., Wang, H. S. and Chung, W. J. “Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation”, *Expert Systems with Applications*, 30, pp. 313–324, 2006.

[22] Byrne, B. M. “Structural Equation Modeling with AMOS: Basic concepts, applications, and programming”, 2<sup>a</sup> ed. (Multivariate Applications Series). New York: Taylor & Francis Group. Traveled, *Structural Equation Modeling*, 11, P. 416, 2010.

[23] Arbuckle, J. L. “AMOS 19 User’s Guide. Chicago: SPSS”, P. 654, 2009.

[24] Dimitrov, D. M. “Comparing groups on latent variables: A structural equation modeling approach”, *A Journal of Prevention, Assessment & Rehabilitation*, 26(4), pp. 429-436, 2006.

[25] Baron, R. M. and Kenny, D. A. “The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations”, *Journal of Personality and Social Psychology*, 51, pp. 1173-1182, 1986.

[26] Mittal, V., Wagner, A. and Kamakura, W. A. “Satisfaction repurchase intent and repurchase behaviour: Investigating the moderating effect of customer characteristics”, *Journal Marketing Research*. 38, pp. 131-142, 2001.

[27] Cooil, B., Timothy, L. K., Lerzan, A. and Micheal, H. “A longitudinal analysis of customer satisfaction and share of wallet: Investigating the moderating effect of customer characteristics”, *Journal Marketing Research*, 71, pp. 67-83, 2007.

[28] Bryant, B. E. and Jaesung, C. “Crossing the threshold”, *Journal Marketing Research*. 8(4), pp. 20-28, 1996.

[29] Assaker, G. and Hallak, R. “Moderating effects of tourists’ novelty seeking tendencies on destination image, visitor satisfaction and short and long term revisit intentions”, *Journal of Travel Research*. 52(5), pp. 600-613, 2013.

[30] Nicolau, J. L. and Mas, F. J. “The influence of distance and prices on the choice of tourist destinations: The moderating effect of motivation”, *Tourism Management*, 27(5), pp. 982-996, 2006.

[31] Perillo, R. D. and Trócolli B. T. “Seria o gênero uma variável discriminante na avaliação da satisfação de clientes? Um estudo com modelagem por equações estruturais”, *rPOT*, 8(1). pp.73-91, 2008.

[32] Linn, L. S., “Burn patients’ evaluation of emergency department care”, *Ann. Emerg. Med.*, 11, pp. 255-9, 1982.