Theoretical and Applied Economics
Volume XVII (2010), No. 6(547), pp. 75-88

Marketing Research on Tourist Consumer Opinions and Behavior in the Center Development Region

Erika KULCSÁR
"Babeş-Bolyai" University, Cluj-Napoca
elerhetetlen@yahoo.com

Abstract. *Logistic regression is a multivariate analysis method of marketing information becoming more and more used because the number of conditions that must be achieved are lower than, for example, in the case of discriminant analysis. The purpose of this paper is to obtain information using logistic regression with respect to the importance of variables in differentiating the classes. The variables included in logistic regression are: the dependent variable is the type of tourists (Romanian or foreign) and the independent variables are as follows: Is this the first time you came to this region? How many days do you plan to stay in this region? Which is the purpose of your staying? In the following year, do you plan to return to this region? This paper also includes a factorial analysis of correspondences, as well as the bivariate analysis using Kruskal-Wallis test in order to simultaneously test the difference between the age of the respondents and their assessment of the organization manner, variety and attractiveness of tours.*

Keywords: variable; Kruskal-Wallis test; logistic regression; Wald test; factorial analysis.

JEL Codes: M31, L83.
REL Codes: 5K, 14G.

1. Introduction

Marketing environment is very complex due to the fact that the very nature of human economic activity is highly complex. This complexity is also found at the level of every form of economic activity. Tourist consumer behavior is very dynamic and continuously changing and out of this consideration it requires constant research.

The bivariate analysis of marketing data can be carried based on contingency tables, which can give us an idea about the existence of differences between population groups, as well as based on statistical tests that allow decisions on the significance of such differences in the total population (Constantin, 2006, p. 135).

The statistical methods of multivariate analysis, according to the relationships that may occur between the dependent variable(s) and the independent variables, can be grouped into two broad categories: the dependencies analysis method (explanatory methods) and interdependence analysis methods (descriptive methods). Dependencies analysis method explains or provides for the development of one or more dependent variables based on three or more independent variables (Lefter, 2004, p. 353).

The multivariate analysis of marketing data was made based on primary data obtained through quantitative research. The quantitative research was conducted between 15.05.2009-17.10.2009. I distributed over 2,000 questionnaires. Questionnaires were distributed to hotels in Braşov, Predeal, Poiana Braşov, Sfântu Gheorghe, Covasna, Miercurea Ciuc, Gheorgheni, Tuşnad, Târgu-Mureş, Sighişoara, Sibiu, Alba Iulia and other localities that have linked this route.

2. Kruskal-Walis test

Non-parametric tests are used for the case where the independent variables are not metrical (Malhotra, 2005, p. 558).

In the event that the comparison of three or more population groups is aimed at, where data are ordinally measured, then for the statistical significance of differences Kruskal-Wallis test is used (Lefter, 2004, p. 266).

Kruskal – Wallis test is the non-parametric counterpart of the one way ANOVA (Marques de Sá, 2007, p. 212).

In the table Ranks (Table 1) the rank means and number of subjects for each group are given (Labar, 2008, p.148).

It can be observed (Table 1) that the mean of ranks is the highest in case of respondents up to age 25 (210.50), followed by respondents over 65 years

(with a mean of ranks 206.75), then subjects between 26-35 years (190.19), followed by tourists who classify between 36-45 years old (173.85), then respondents aged 46-55 years (158.06) and finally respondents between 56-65 years (with a mean of ranks 206.75).

Table 1

Mean ranks for Kruskal-Wallis test – the organization manner

Ranks

	Age	N	Mean Rank
Organization manner	up to 25 years	24	210.50
	between 26-35 years	118	190.19
	between 36-45 years	88	173.85
	between 46-55 years	69	158.06
	between 56-65 years	46	154.93
	more than 65 years	8	206.75
	Total	353	

Table Statistical Test (Table 2) presents the Kruskal Walis value (Chi-Square), the degrees of freedom (df) and the significance threshold (Asymp. Sig.) (Labar, 2008, p.148).

In this case df = 5 (df = k-1, where k is the number of groups). The critical value $\chi^2_{20,05;5} = 11{,}071$.

H value $= \chi^2_{calc}$ in the analysis table is equal to 10.673, hence null hypothesis H_0 is accepted.

Table 2

Values calculated with the Kruskal-Walis test

Test Statistics[a,b]

	Organization manner
Chi-Square	10.673
df	5
Asymp. Sig.	.058

a. Kruskal Wallis Test

b. Grouping Variable: Age

Table 3 contains the size of each group in the sample and the mean ranks corresponding to the groups.

Table 3

The mean ranks for Kruskal-Wallis test - the variety of tours

Ranks

	Age	N	Mean Rank
Variety of tours	up to 25 years	24	215.25
	between 26-35 years	118	182.47
	between 36-45 years	90	193.43
	between 46-55 years	69	151.04
	between 56-65 years	44	158.23
	more than 65 years	8	123.88
	Total	353	

Table 4

The calculated values for Kruskal-Walis test

Test Statistics[a,b]

	Variety of tours
Chi-Square	15.371
df	5
Asymp. Sig.	.009

a. Kruskal Wallis Test

b. Grouping Variable: Age

In this case H = 15.371> $\chi^2_{20,05;5} = 11{,}071$ (Table 4), which means that the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted.

Analyzing the data in Table 5 it may be noted that the mean of the ranks is higher for the tourists up to 25 years, and the lowest for respondents over 65 (111.13).

Table 5

Ranks average for Kruskal-Wallis test - the attractiveness of tours

Ranks

	Age	N	Mean Rank
Attractiveness of tours	up to 25 years	24	224.17
	between 26-35 years	118	185.25
	between 36-45 years	90	201.41
	between 46-55 years	69	146.83
	between 56-65 years	42	128.31
	more than 65 years	8	111.13
	Total	351	

From Table no. it can be seen that value H = 34.952> $\chi^2_{20,05;5} = 11{,}071$ (table no). Conclusion: the alternative hypothesis H_1 is accepted.

Table 6

Values calculated wit Kruskal-Walis test

Test Statistics[a,b]

	Attractiveness of tours
Chi-Square	34.952
df	5
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: Age

3. Logistic regression

The logistic regression(1) models the relationship between a set of independent variables (categorical, continuous) and a Y dichotomous dependent variable (nominal, binary). Unlike classical regression, here one or more predictive/explanatory variables (independent variables) may be categorical, compulsorily the dependent variable, so with this new model it is about the non-numeric qualitative (categorical) nature of some of its variables(2).

The logistic regression model shall estimate the proportion of subjects which will have the same interest characteristic for the statistical status, or equivalently, the probability that a person among the population have a certain characteristic.

The results obtained by using SPSS program are presented below. The first table provides general information: number of valid observations, number of non-valid observations and sample size (Table 7).

Table 7

General data regarding observations

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	410	100.0
	Missing Cases	0	.0
	Total	410	100.0
Unselected Cases		0	.0
Total		410	100.0

a. If weight is in effect, see classification table for the total number of cases.

For independent categorical variables(1) a recoding occurs by taking into account the reference class: this is recoded 0 (Table 8).

Table 8

Dependent variable recoding

Dependent Variable Encoding

Original Value	Internal Value
Romanian	0
foreign	1

The ninth table presents the frequencies related to the categorized dependent variable.

Table 9

Frequencies of the dependent variable

Categorical Variables Codings

		Frequency	Parameter coding (1)
Is this the first time when you come to this region?	No	222	1.000
	Yes	188	.000

The output differs (1) in structure by the method of variables selection, but it contains a first block of information that relates to the simple model (only with the term constant). Note the structure: classification, variables in equation, candidate variables.

Block 0: Beginning Block

Table 10

Classification table

Observed			Predicted		
			T_tourist		Percentage Correct
			Romanian	foreign	
Step 0	T_tourist	Romanian	242	0	100.0
		foreign	168	0	.0
	Overall Percentage				59.0

a. Constant is included in the model.

b. The cut value is .500

The Classification table (1) is made by considering the classification probabilities predicted by the current model for each observation by the principle that the OR> 1 ranks the observation in the 1st codified group.

The first phase of the analysis indicates that 242 of the tourists belong to the first group, i.e. 59.5% of the sample, while of 168 the tourists belong to the 2nd group.

The percentage of tourists that belong to the first group is also important for the fact that if I was chosen randomly among interviewed tourists and I was to foresee that they belong to the first group, then I would have been right for 59.0% (Table 10).

Next the tables obtained before the analysis phase are presented which is why Table 11 contains only the constant.

In order to see whether the coefficients are significant or not, the Wald statistics is used for the estimation of the parameters, which is significant. (0.000).

Wald-test is the square of z (standard coefficient/error). For larger samples the p's for Wald and the z are equal.

Table 11

Parameter estimation

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.365	.100	13.209	1	.000	.694

The table below Variables not in the Equation (Table 12), as the title suggests, shows that the independent variables have not yet been introduced in the analysis so the program shows no association between these independent variables, but it provides information on the effect of each independent variable separately. Thus, he 3rd variable (Is this the first time you came to this region?), variable 6 (How many days do you plan to stay in this region?) and variable 7 (Which is the purpose of your staying?) indicate a significant effect, while variable 32 (In the following year, do you plan to return to this region?) does not have a significant effect.

Table 12

The effect of independent variables

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	var3(1)	11.691	1	.001
		var6	11.912	1	.001
		var7	7.177	1	.007
		var32	1.592	1	.207
	Overall Statistics		34.597	4	.000

The second part of the analysis (Block 1: Method = Enter) shows final results.

The four independent variables were entered together in the analysis; this actually means the Enter process.

Based on the results obtained in Table 13, it can be ascertained that the model (following the introduction of the independent variables) is significant.

Table 13

Omnibus Test

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	37.632	4	.000
	Block	37.632	4	.000
	Model	37.632	4	.000

The results of Table 14 - Summary Model indicates the percentage by which the variance of the dependent variable is explained by the combination of independent variables. In this respect the system provides two indicators: Cox & Snell R Square and Nagelkerke R Square.

According to the first indicator, the four independent variables explain with a 8.8% rate while the last 11.8% percentage, the variance of the dependent variable. (According to specialists the indicator Negelkere R Square should be considered as the Cox & Snell R Square indicator underestimate the true value).

Table 14

Model results

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	517.319[a]	.088	.118

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

According to the results in Table 15, 81.0% is the proportion of those subjects who were Romanian tourists and they have been correctly categorized by the model, on the other 40.5% is the proportion of those correctly categorized in the case of foreign tourists.

This means that eventually a percentage of 64.4% of cases was correctly classified by comparing it with the expected success of categorization (59.5%) (Table 4), it can be ascertained that the independent variables contributed with a 4.9% percentage to the correct categorization of thee dependency.

Table 15

Classification table

Observed			Predicted		
			T_tourist		Percentage Correct
			Romanian	foreign	
Step 1	T_tourist	Romanian	196	46	81.0
		foreign	100	68	40.5
	Overall Percentage				64.4

a. The cut value is .500

The last table in Block 1: Method = Enter as compared to table in Block 0: Beginning block includes the combined effect of all independent variables (Table 16).

In the table referring to the variables in the model it is reported:

- the B coefficients,
- Exp (B) with the interpretation, given theoretically, that it is the change of OR (chances rapport) of the dependent variable to the modification with a unit of the independent variable, so Exp B ≈ 1 for insignificant variables.
- Information associated to the Wald significance test of each coefficient [(1)].

Table 16

The combined effect of independent variables

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	5.0% C.I.for EXP(B	
								Lower	Upper
Step 1[a]	var3(1)	-.671	.215	9.766	1	.002	.511	.336	.779
	var6	-.112	.029	15.229	1	.000	.894	.845	.946
	var7	-.168	.057	8.650	1	.003	.845	.755	.945
	var32	-.329	.208	2.509	1	.113	.719	.479	1.081
	Constan	1.521	.396	14.725	1	.000	4.575		

a. Variable(s) entered on step 1: var3, var6, var7, var32.

It may be noted that Var3 (Is this the first time that you come to this region?) Var6 (How many days do you plan to stay in this region?) Var7 (Which is the purpose of the staying?) still have a significant effect.

4. Factor analysis of correspondences

The factor analysis of correspondences aims at describing the relationships between two nominal variables. Although it is not a multivariate analysis method, but a bivariate one, the relationships between variables modalities can be described through the factorial analysis, which can not be surprised by the test.

The primary objective of the analysis of correspondence is the simultaneous study of the rows and columns of a contingency table in order to highlight the links and correspondences between the two assemblies (Lefter, 2004, p. 421).

A first set of results is shown in the table of correspondences between the two variables: In what country do you live? and Would you recommend the hotel where you stayed to your friends and acquaintances? The contingency table, showing the distribution of Romanian and foreign respondents according to the two variables, is as follows (Table 17).

Table 17

Correspondence table

In what country do you live?	Would you recommend the hotel where you stayed to your friends and acquaintances?			
	no	yes	I do not know	Active Margin
Romania	16	169	57	242
SUA	0	20	2	22
Great Britain	0	20	0	20
Germany	0	30	2	32
France	0	8	0	8
Holland	0	0	2	2
Spain	0	8	2	10
Australia	0	2	0	2
Switzerland	2	8	2	12
Italy	0	4	0	4
Israel	0	2	0	2
India	0	4	0	4
Austria	0	2	0	2
Hungary	0	40	6	46
China	0	2	0	2
Active Margin	18	319	73	410

In the correspondence table, the totals corresponding to the rows and columns are called Active Margin (Constantin, 2006, p. 261).

From the crossed analysis of frequencies corresponding to the categories of the two variables some differences can be noticed between the frequencies corresponding to the categories of responses, in that their distributions are not uniform. But their interpretation is quite difficult.

The sample includes 410 tourists. I wanted to analyze the relationship between the data in the table above in order to identify correspondences between them.

Using SPSS software I obtained the following results:

Table 18

Statistics on factorial axes and their importance

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	.307	.094			.839	.839	.033	.140
2	.135	.018			.161	1.000	.055	
Total		.112	46.104	.017[a]	1.000	1.000		

a. 28 degrees of freedom

In the above table (Table 18) the first column indicates the existence of the two factorial axes (dimensions).

In the Singular Value column are the values that represent the importance of the two dimensions, while in the Inertia column there are the squares of the individual values, representing a total dispersion of the points consisting of individuals belonging to each category (Constantin, 2006, p. 261).

The third column shows a measure of the total variance of the points taking into account their marginal rate, and their values λ_1 and λ_2, respectively. It is the square pf the singular value (Lefter, 2004, p. 424).

Differences among frequencies of the categories and relationships between the two variables are highlighted by Chi Square value, leading to acceptance of alternative hypothesis H_1 of the existence of a connection between the two variables, whereas the minimum level of significance that can be accepted for this hypothesis is less than 0.05 (Sig. = 0.017).

In the columns Proportion of Inertia there are the proportions between the two factors that explain the scattering of data. It is noted that the first component explains the scattering with a 83.9% rate, the second at 16.1% rate.

This can be also viewed in the graphical representation in Figure 1, where there is a greater proximity to the first axis categories.

From the chart below it can be observed that the first factorial axis, namely the 1st factor, is explained by the yes answer variant, the second factorial axis, respectively the second factor, is explained by the do not know or no answer variants. Differences between the three categories of answers are obvious.

Graphic expression of the above results is as follows:

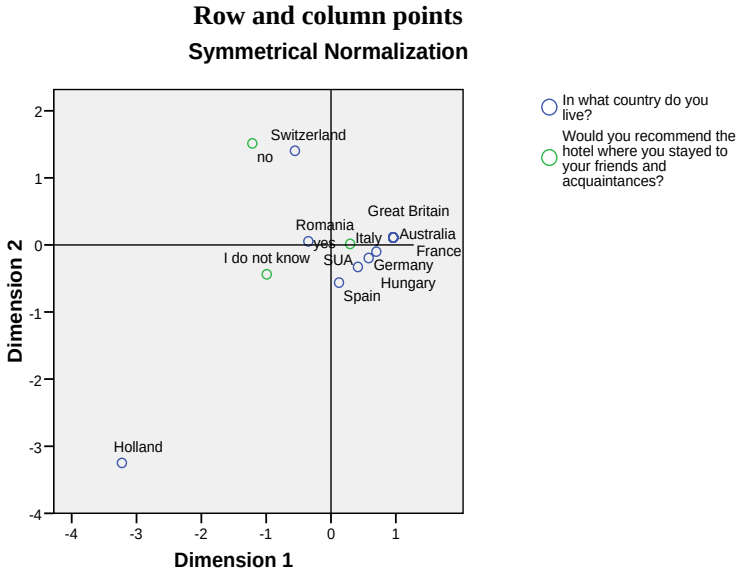

Figure 1. *Graphical representation of the correspondences between the categories of response*

5. Conclusion

Following the analysis of Kruskal Wallis test I found that:

- Between the six groups of population (ages of respondents) there is no difference in terms of assessing the organization of tours.
- Between the six groups of population (ages of respondents) there is a difference in the assessment by the tourists of the variety of excursions.
- There are differences between the six groups of population (ages of respondents) in assessing the attractiveness of tours.

Development and modernization of services (...) is one of the decisive factors in increasing the competitiveness of the offer on the international market. Developments in recovering Romania's tourism potentials are still modest (Nita, Nita, 2008, p. 198).

For these reasons I believe that companies operating on the tourism market in the Center Development Region should concentrate their efforts, for the period to come, on strategic alternatives and they should understand that if they don't have something special to offer they have no chance to assert themselves. A person generally wants to discover new places, new entourages, new people, other lifestyles, other civilizations and cultures. He seeks exotic, unusual, miracle, wellness. From this point of view companies should adapt the marketing strategies to the needs and expectations of each segment of tourists since as outlined by the bivariate analysis, there is a differences between tourists in terms of group age and the assessment of the variety and attractiveness of tours. But in this analysis I included both Romanian tourists and foreign tourists and it is well known, the consumer behavior, even if it falls in the same age category, is different not only from one country to another but from one region to another. For example in Romania, most tourists in balneal resorts are pensioners and people over 45-50 years, who wish to follow treatment procedures. In West spas are dedicated especially to prevention and restoration and not treatment. The current profile of balneoclimateric tourism will need to bear changes because the current model, of treatment of various diseases, is not in accordance with the European model, which emphasizes prevention and recovery.(3) In my opinion companies operating in the tourist market of the Centre Development Region should select a particular segment to be very well served.

As a final conclusion of the logistic regression analysis I found that, forecasting the days spent, the purpose of the staying and whether or not they are for first in this region of the country (Centre Development Region) differentiate Romanian tourists from the foreign tourists. Regarding the variable

In the next years do you plan to return to the region? it does not differentiate Romanian tourists from foreign tourists.

And following the factorial analysis of correspondences there are connections between the two variables, namely the variables "in which country the respondent reside" and "if they would recommend the hotel they have stayed in to friends and acquaintances".

As a final conclusion multivariate analysis of marketing data can identify differences between Romanian and foreigners consumer behavior which highlights once again that it is extremely important for companies operating on the tourist market in the Center Development Region to understand the importance of market segmentation because market segmentation is one of the main sources of success on the market.

Notes

(1) see http://thor.info.uaic.ro/~val/statistica/StatWork_10.pdf**.**
(2) see http://inf.ucv.ro/~gorunescu/courses/curs/7.pdf**.**
(3) see http://www.newschannel.ro/stiri/39/4239/Aderarea-va-creste-ponderea-turismului.html, newschannel.ro, 2006.

References

Constantin, C. (2006). *Sisteme informatice de marketing. Analiza şi prelucrarea de marketing. Aplicații în SPSS*, Editura Infomarket, Braşov

Labăr, A. V (2008). *SPSS pentru ştiințele educației*, Editura Polirom, Iaşi

Lefter, C. (2004). *Cercetarea de marketing* Editura Infomarket, Braşov

Malhotra, K.N, *Cercetarea de marketing*, Editura Akadémiai Kiadó, Budapesta

Marques de Sá J. P. (2007). *Applied Statistics. Using SPSS, STATISTICA, MATLAB and R*, Editura Springer Berlin Heidelberg, New York

Niță, I., Niță, C. (2008). *Piața turistică a României*, Editura Economică, Bucureşti