THE BIVARIATE ANALYSIS BETWEEN THE VARIABLES THAT DEFINE THE INVESTIGATED TOURIST POPULATION IN THE CENTER DEVELOPMENT REGION AND THE OTHER IMPORTANT TOURIST VARIABLES

Assist. Ph.D Kulcsár Erika
University "Babeş-Bolyai"
Faculty of Economics and Business Administrations
Cluj Napoca, Romania

Abstract: In the marketing research the instances when we need to examine the relationship between two variables are frequent. Knowing the relationship between the two variables involves the use of tests that can be parametric and nonparametric. This paper includes the non-parametric tests used in the bivariate analysis. The used tests are: chi square test, the Mann-Whitney test, the Kolmogorov-Smirnov test. Besides these tests, this paper also includes measuring the association between nominal variables using the C contingency coefficient and Cramer's V coefficient and between two metric variables, using the Pearson's linear correlation coefficient.

JEL classification: M31, L83

Key words: variables, χ^2 test , U test, Kolmogorov-Smirnov test, Cramer's V contingency coefficient, Pearson's linear correlation coefficient,.

Introduction

The relation between two variables can be measured with the same type of scale or, each with a different type of scale. In this case, practically, I plan to analyze the link between two questions. Actually I want to know whether the answers given by the subjects to a question, is correlated or associated with the answers given to another question. [6]

The quantitative marketing research took place during 15.05.2009-17.10.2009. Over 2,000 questionnaires have been distributed. 410 filled in questionnaires have been included in this research. The questionnaires were filled in by both foreign tourists and Romanian tourists 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.

Non-parametric test used in the bivariate analysis χ^2 test (chi-square)

The χ^2 test is used to test the statistical significance of the frequencies distribution, as absolute or percentage values from two variables measured with nominal scale, having a given number of ways, for a sample of predetermined size.

[6] Constantin L. Marketing Research, Infomarket Printing House, Braşov, 2004. p. 249

Also, the chi square test can also be used to measure the differences between two or more independent groups when a certain variable is taken into account. The responses given by the subjects must be introduced in a contingency table.[7]

Next I will try to highlight the link between the age of the respondents and whether they came for the first time or not in the Centre Development Region.

Table no. 1. The contingency table

Is this the first time when you come to this region? * Age Crosstabulation

			Age						
			till 25 years old	between 26-35 years old	between 36-45 years old	between 46-55 years old	between 56-65 years old	more than 65 years old	Total
Is this the first time wh	No	Count	14	44	80	48	28	8	222
you come to this regio		Expected Cour	14.1	65.0	57.4	52.0	28.2	5.4	222.0
	Yes	Count	12	76	26	48	24	2	188
		Expected Cour	11.9	55.0	48.6	44.0	23.8	4.6	188.0
Total		Count	26	120	106	96	52	10	410
		Expected Cour	26.0	120.0	106.0	96.0	52.0	10.0	410.0

Within table no. 1 the distribution of the tourists who came for the first time and those who came before in the region according to the six age categories can be noticed.

There are differences between the observed frequencies and the expected frequencies in all subgroups formed by crossing the two variables.

Table no. 2. χ^2 test results

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	37.543[a]	5	.000
Likelihood Ratio	38.975	5	.000
Linear-by-Linear Association	3.986	1	.046
N of Valid Cases	410		

a. 1 cells (8.3%) have expected count less than 5. The minimum expected count is 4.59.

χ^2_{calc} value is 37.543, it will be compared with $\chi^2_{0,05;5}$ = 11.071 (table no. 2). Since $\chi^2_{calc} > \chi^2_{0,05;5}$, H_0 is rejected in the sense that we can guarantee with a 95% probability that at the level of the total population, there will be differences between the expected and observed frequencies. In conclusion, there is a correlation between the age of the tourists and whether they had been before or not in the Center Development Region.

Analyzing the level of significance I reached the same conclusion. Since it is less than 0.05, according to which there is a difference between the two variables, H_1 is accepted.

[7] Constantin L. Marketing Research, Infomarket Printing House, Braşov, 2004. p. 250

Mann-Whitney test (U test)

The U test allows comparison of the differences in the case when the respective populations are not normally distributed and they are not equal in terms of variances. The specific of the test is that to quantitative values corresponding to two independent groups, ranks are associated which brings to study the distribution of the ranks between the two samples. [8] In this part of the essay I will test the existence of a significant difference between the two sex categories with respect to the importance of the level of charged tariffs in choosing a hotel.

Table no. 3 The sum of the ranks for each group

Ranks

	Gender	N	Mean Rank	Sum of Ranks
Level of charged tariffs	Male	230	214.77	49397.00
	Female	180	193.66	34858.00
	Total	410		

After applying the test, I obtained table no. 3 which contains the size of each group in the sample, the amount of ranks and average ranks. In another table provided by SPSS, the value U_{obs} is given, together with other sizes used to deciding in the case of the U test (table no. 4)

Table no. 4 Values corresponding to the U test

Test Statistics [a]

	Level of charged tariffs
Mann-Whitney U	18568.000
Wilcoxon W	34858.000
Z	-1.861
Asymp. Sig. (2-tailed)	.063

a. Grouping Variable: Gender

In both cases the sub-samples are large, both exceeding 30 people ($n_1 = 230$ and $n_2 = 180$). The decision was taken based on the calculated size z_U=-1.861, which was compared to the value in the table for a significance level α = 0.05. It is noted that z_U=-1.861 >-z_U=-1.96. Therefore we accept the null hypothesis, according to which we can not guarantee with a 95% probability that, between male and female, there are differences in terms of the importance given to the level of prices charged, in choosing a hotel. I could have taken this decision also based on the level of significance "Asymp. Sig. (2-tailed), which is 0.063, higher than the 0.05 value, which leads also to the acceptance of the null hypothesis.

Kruskal-Walis test

In the event the comparison of three or more groups of the population is aimed at, where data are measured ordinal, then the Kruskal-Walis test is used for the statistical significance of the differences.[9]

[8] Constantin L. Marketing Research, Infomarket Printing House, Braşov, 2004. p. 259

[9] Cristinel C. - Marketing information system. Marketing data analysis and processing. Applications in SPSS, Publisher Infomarket, Braşov, 2006, p. 172

I will test the existence of significant differences between 6 groups of population formed according to the age of tourists regarding the importance of prices charged to choose a hotel.

The SPSS system has provided a table containing the size of each group at the level of the sample and the mean ranks corresponding to the groups (table no. 5).

Table no. 5 The mean ranks in the case of the Kruskal-Walis test

Ranks

	Age	N	Mean Rank
Level of charged tariffs	till 25 years old	26	217.50
	between 26-35 years old	120	191.15
	between 36-45 years old	106	237.93
	between 46-55 years old	96	174.27
	between 56-65 years old	52	211.96
	more than 65 years old	10	268.90
	Total	410	

Table no. 6 The calculated values in the case of the Kruskal-Wallis test

Test Statistics [a,b]

	Level of charged tariffs
Chi-Square	21.220
df	5
Asymp. Sig.	.001

a. Kruskal Wallis Test

b. Grouping Variable: Age

The value H = χ^2_{calc} of the analysis table is equal to 21.220 (table no. 6).Since H = 21,220 > 11.071, which means that I accept the alternative hypothesis. Therefore, between the 6 groups of the population there are differences in terms of the importance given to the levels of prices charged in choosing a hotel.

Kolgomorov-Smirnov test

The test is used to test the differences between two independent samples, the same as with Mann-Whitney test. Test analyses the existence of significant differences between the two groups in terms of evaluations or attitudes towards a product, service or marketing phenomenon, based on a specific criterion. [10]

Based on Kolmogorov-Smirnov, I shall test if the views on the statement "The gentleness of the staff can make this hotel to become one of the favorite places of tourists" are different depending on the gender of the respondents.

Out of the 180 female respondents 43.3% have chosen the "totally agree", 4.4% opted for "totally disagree". 39.1% of male opted for the variant “totally agree" (table no. 7).

[10] Cristinel C. - Marketing information system. Marketing data analysis and processing. Applications in SPSS, Publisher Infomarket, Braşov, 2006, p. 166

Table no.7. Relative frequencies n percentages and columns

			Gender		
			Male	Female	Total
The gentleness of the staff can make this hotel to become one of the favorite places of tourists	totally disagree	Count	6	8	14
		% within Gender	2.6%	4.4%	3.4%
	disagree	Count	8	8	16
		% within Gender	3.5%	4.4%	3.9%
	neither agree nor disagree	Count	58	22	80
		% within Gender	25.2%	12.2%	19.5%
	agree	Count	68	64	132
		% within Gender	29.6%	35.6%	32.2%
	totally agree	Count	90	78	168
		% within Gender	39.1%	43.3%	41.0%
Total		Count	230	180	410
		% within Gender	100.0%	100.0%	100.0%

Table no. 8 Distribution of responses on groups formed according to the gender

Frequencies

	Gender	N
The gentleness of the staff can make this hotel to become one of the favorite places of tourists	Male	230
	Female	180
	Total	410

Out of the total of 410 tourists who have expressed their opinion on this question, 180 are female and 230 male (table no. 8)

Table no. 9. The values calculated for the Kolmogorov-Smirnov test

Test Statistics[a]

		The gentleness of the staff can make this hotel to become one of the favorite places of tourists
Most Extreme Differences	Absolute	.102
	Positive	.102
	Negative	-.028
Kolmogorov-Smirnov Z		1.024
Asymp. Sig. (2-tailed)		.245

a. Grouping Variable: Gender

The maximum difference between the cumulated frequencies is 0.102. D_{calc} = 0.102 (table no. 9) I compared the value of for D_{α} $n_1 = 230, n_2 = 180$. I calculated

this value according to the formula: $D_\alpha = 1{,}36\sqrt{\frac{n_1+n_2}{n_1 \times n_2}} = 1{,}36\sqrt{\frac{230+180}{230\times 180}} = 0{,}135.$

Since $D_{calc} < D_\alpha$, the alternative hypothesis is accepted.

Conclusion: Between the two groups there are no differences in terms of their opinions regarding the gentleness of staff.

Measuring the association between nominal variables

The χ^2 test only proved me the existence of links between the nominal variables analyzed without indicating their intensity. In order to measure the association between nominal variables I used the C contingency coefficient and Cramer's V coefficient.

Next I will try to highlight the intensity of the relationship between age and whether respondents had been to this region of the country before or not. This relationship had been previously examined, through the χ^2 test, the result being the fact that we can accept with a 95% probability that the two variables are related.

$$C = \sqrt{\frac{\chi^2_{calc}}{\chi^2_{calc}+n}} = \sqrt{\frac{37.543}{37.543+410}} = 0.290$$

From table no. 2 I extracted the value χ^2_{calc} which is 37.543 and with its help I can calculate the coefficients that measure the association between the two variables.

Independent variable is age, which has 6 categories. Therefore the basic value to which $C_{\max}$ is calculated is k = 6. $C_{\max} = \sqrt{\frac{k-1}{k}} = \sqrt{\frac{6-1}{6}} = 0.91$

Next I shall calculate also the value of the Cramer's V coefficient.

$$V = \sqrt{\frac{\chi^2_{calc}}{n\times t}} = \sqrt{\frac{37.543}{410\times 1}} = 0.303$$

In table no. 10 the measuring coefficients of the association between the nominal variables provided by the SPSS system are presented.

Table no. 10 The measuring coefficients of the association between the nominal variables

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.303	.000
	Cramer's V	.303	.000
	Contingency Coefficient	.290	.000
N of Valid Cases		410	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

The calculated coefficients indicate an average association between the two variables.

Measuring the association between two metric variables

The intensity of the relationships between two variables metrically measured is measured based on the Pearson's linear correlation coefficient. It is based on the co-variation between the two variables, being a parametric coefficient since it is calculated as the mean of the two variables.[11]

Table no. 11 Pearson's linear correlation coefficient

Correlations

		How many days did you plan to stay in this region?	Estimation of the expenes incurred in the Center Development Region (including the residence country - EUR-)
How many days did you plan to stay in this region?	Pearson Correlation	1	.017
	Sig. (2-tailed)		.729
	N	410	410
Estimation of the expenes incurred in the Center Development Region (including the residence country - EUR-)	Pearson Correlation	.017	1
	Sig. (2-tailed)	.729	
	N	410	410

The correlation coefficient between the two variables is r = 0.017 (table no 11), this indicating that there is no significant link between the expenditure incurred in the Centre Development Region and the number of days planned for this region. For the test I calculated the critical report: $RC = \frac{r\sqrt{n-2}}{n-r^2} = \frac{0{,}017 \cdot \sqrt{410-2}}{\sqrt{1-0{,}017^2}} = 0.343$

RC=0,343<1.96, which means that the two variables are not correlated. This decision can be taken also based on the level of significance which is higher than 0.05.

By processing the primary data I obtained the following results:

- There is a link between the age of respondents and whether they had been or not in the Center Development Region before. Most of those respondents who had been in this part of the country before fall within the age group 26-35 years category followed by tourists with ages between 46-55 years.
- The results obtained based on the U test surprised me, namely that there is no difference between male and female regarding the importance given to the level of prices charged, in choosing a hotel. I was very convinced that there is a connection between the two groups and the importance given to charges of the hotel. But on the one hand, these results underline the fact that the behavior of services consumer differs from that of the products consumers, others than services.

[11] Cristinel C. - Marketing information system. Marketing data analysis and processing. Applications in SPSS, Publisher Infomarket, Braşov, 2006, p. 203

- Between the six age groups there are differences in the importance given to the levels of charged in choosing a hotel which actually explains the existence on the tourist market of different categories of hotels.
- With respect to the statement "Gentleness of staff makes this hotel to become one of the favorite places of tourists" , the opinion of female respondents and that of male respondents was similar, i.e. there is no difference between the two groups and their evaluations on the gentleness of staff hotel.
- Although I expected that after the application of the Pearson's linear correlation coefficient to discover a relationship between the costs incurred in the Center Development Region and the expected number of days in the area, following the interpretation of the results obtained I found that there is no significant relation between the two mentioned variables.

As a final conclusion of this paper I can mention that the behavior of the tourist services consumer requires further research.

REFERENCES

1. Cristinel, C. Marketing information system. Marketing data analysis and processing. Applications in SPSS, Infomarket Press, Braşov, 2006.
2. Constantin, L. Marketing Research, Infomarket Printing House, Braşov, 2004.