MULTIPLE REGRESSION ANALYSIS OF MAIN ECONOMIC INDICATORS IN TOURISM

Assistant Ph.D. **Erika KULCSÁR**
"Babeş Bolyai" University of Cluj Napoca, Romania

Abstract

This paper analysis the measure between GDP dependent variable in the sector of hotels and restaurants and the following independent variables: overnight stays *in the establishments of touristic reception, arrivals in the establishments of touristic reception and investments in hotels and restaurants sector in the period of analysis 1995-2007. With the multiple regression analysis I found that investments and tourist arrivals are significant predictors for the GDP dependent variable. Based on these results, I identified those components of the marketing mix, which in my opinion require investment, which could contribute to the positive development of tourist arrivals in the establishments of touristic reception.*

Key words: *coefficient of determinations -* R^2*, analysis of variance, Student test, multivariate analysis of dependencies*

JEL classification: *D12, L83*

1. INTRODUCTION

Regression analysis generally belongs to the multivariate methods, while being an explanatory method of analysis.

The regression analysis describes through a model the relationship between a dependent variable and one or more independent variables.

The GDP achieved in the hotels and restaurants sector may depends on the investments in hotels and restaurants, the arrivals of tourist arrivals in the establishments of touristic reception, the overnight stays in the establishments of touristic reception. The main purpose of this analysis is to determine to what extent GDP is affected by the three independent variables and which are the measures that should be taken based on the results - obtained by using the SPSS program - in tourism marketing field.

Objectives of the study:

1. Analysis of variance (ANOVA) regarding multiple regression

2. To achieve a statistical test (Student test) for each independent variable, in order to know which regression coefficient can be 0 and which can not be.

3. To determine the intensity of the interdependence connection between independent variables

2. MULTIPLE REGRESSION

The multiple regression analysis studies the simultaneous emotions that two or more independent variables may have over one dependent variable (Lefter, 2004, p.364)
.

Table 1 - Evolution of the main economic indicators in tourism in the period 1995-2007

Years	GDP obtained in the sector of hotels, restaurants (million RON updated prices)*	Investments in hotels and restaurants (million RON in updated prices)*	Tourists arrivals in the establishments of touristic reception (total, thousands of tourists)	Overnight stays the establishments of touristic reception (total, thousands)
1995	8461.5	1389.0	7070	24111
1996	11924.2	1040.7	6595	21838
1997	9375.7	627.3	5727	19611
1998	7758.8	581.4	5552	19183
1999	6072.9	526.3	5109	17670
2000	5491.7	316.7	4920	17647
2001	5237.2	584.8	4875	18122
2002	4586.9	557.0	4847	17277
2003	4556.1	704.4	5057	17845
2004	4850.4	956.9	5639	18501
2005	5872.2	1080.6	5805	18373
2006	7060.5	1332.0	6216	18992
2007	8402.4	1600.9	6972	20593

Source: Romanian Statistical Year Book, 2008, pp. 477, 517, 826-827
* upgraded values (2007 reference period)

Using the SPSS programs kit in the case of multiple regressions I have come to the following results:

Table 2 - Regression Coefficients:

Model	Indicators	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-16206.351	3622.650		-4.474	.002
	Investments in hotels and restaurants (million RON in current prices)	-8.173	2.167	-1.463	-3.771	.004
	Tourists arrivals in the establishments of touristic reception (total, thousand tourists)	6.584	1.749	2.340	3.765	. 004
	Overnight stays the establishments of touristic reception (total, thousands)	-.389	. 414	-.344	-.939	.372

Dependent Variable: The Gross Domestic Product obtained in the hotels and restaurants sector (millions RON, current prices)

Based on nonstandard coefficient we get the regression equation:

$$\hat{y} = -16206.351 - 8.173\, x_1 + 6.584\, x_2 - 0.389\, x_3$$

(1)

where

x_1 - represents the investments in hotels and restaurants (millions RON, current prices)

x_2 - represents tourists arrivals in the establishments of touristic reception (total, thousands of tourists)

x_3 - represents overnight stays the establishments of touristic reception (total, thousands)

Table 3 - Estimation of Standard Deviation
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.916(a)	.840	.786	1019.0549

Predictors: (Constant), Overnights stays the establishments of touristic reception (total, thousands), Investments in hotels and restaurants (millions RON, current prices), Tourists arrivals in the establishments of touristic reception (total, thousands of tourists)

The coefficient of determination R^2 indicating the percent of how much of the total variance is

explained by the independent variable is 84% (*Table 3*).

The analysis of variance for multiple regressions will be made starting from the following results:

Table 4 - Variation analysis
ANOVA (b)

Model		Sum of Squares	df	Mean Square	F	Sig.
	Regression	49017946.282	3	16339315.427	15.734	.001(a)
	Residual	9346256.147	9	1038472.905		
	Total	58364202.429	12			

a. Predictors: (Constant), Overnights stays the establishments of touristic reception (total, thousands), Investments in hotels and restaurants (millions RON, current prices), Tourists arrivals in the establishments of touristic reception (total, thousands of tourists)

b. Dependent variable: The Gross Domestic Product obtained in the hotels and restaurants sector (millions RON, current prices)

The result is that most part of the total variance is generated by the regression equation. If we would make the report between the variance owed to regression (SSR) and the total variance (SST) we would get the coefficient of determination R^2, highlighted in the previous table (*Table 3*).

$$R^2 = \frac{SSR}{SST} = \frac{49017946.282}{58364202.429} = 0.840 \ (2)$$

This means that 84% of the total variance is explained by the regression model while 16% by unknown factors, that variance representing the error (SSE).

Starting from this error the multiple standard error of estimation is gained based on the following relation:

$$S_{y^3} = \sqrt{\frac{SSE}{n-(k+1)}} = \sqrt{\frac{9346256.147}{13-(3+1)}} = 11019.0549$$

(3)

In order to test the validity of multiple regression model a global test must be used that researches whether all the independent variables have regression coefficients equal with 0, or in other words if the explained variance is not due to a random. The regression coefficients of the sample have as correspondent at the population level the regression coefficient marked as $\beta_1, \beta_2, \beta_{3,...}$

The alternative and null hypotheses are formulated as follows:

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_1 = \text{not all } \beta \text{ coefficients are equal to zero}$$

In order to test the null hypothesis we turn to F test that requires an analysis of the variance identified in the ANOVA table above.

From the data in the previous table (*Table 4*) it can be ascertained that the value of the calculated F is 15.734 for the variance generated by the regression. The critical value of F, at the significance level of 0.05 with 3 degrees of freedom at numerator and 9 at denominator, is 3.86. By comparing the values of F it results that it is compulsory to accept the alternative hypothesis, meaning the fact that not all the regression coefficients are equal to zero. This means that a significant influence of multiple regression model occurs over dependent variables.

The issue that arises now is to know which regression coefficients may be zero and which not. It is imposed therefore to achieve an individual evaluation of regression coefficients. It is therefore compulsory to make an assessment the realization of a statistical test for each under the conditions where the null hypothesis implies that each coefficient β is equal to zero and the alternative hypothesis states that it is different from zero.

The test used is the Student test, respectively t with n-(k+1) degree of freedom. For each of the three variables, from the SPSS results, we get the calculated t values (*Table 2*). These are: -3.771 for investments in hotels and restaurants, 3.765 for tourists' arrivals in the establishments of touristic reception, -0.939 for overnight stays the establishments of touristic reception.

In order to define the decision rule concerning the null hypothesis, the calculated t values will be compared with the critical value of t at a significance level of 0.05 in the case of a two-tailed test, with 13-(3+1), meaning with 9 degrees of freedom. This value is ± 2.262. The result is:

In the cases when investments calculated *t* (-3.771) is lower than critical *t* (-2.262). The level of significance indicated by the test 0,004 is lower than the chosen level of significance of 0.05. Therefore the null hypothesis is rejected and it is accepted that β_1 is

different from zero.

• In the case of the variable "tourists arrivals in the establishments of touristic reception" the calculated t (3. 765) is higher than critical t (2.262). The level of significance indicated by the test of 0.004 is lower than the chosen level of significance 0.05. Therefore the null hypothesis is rejected and it is accepted that β_2 is different from zero.

• In the case of overnight stays the establishments of touristic reception, the calculated t (-0.939) is higher than critical t (-2.262). Likewise level of significance indicated by the test of 0.372 is

higher than the chosen level of significance 0.05. In this case the null hypothesis is accepted and it is considered that β_3 is equal to zero. Therefore, it is considered that the variable "overnight stays in the establishments of touristic reception" is not a significant predicator for the dependent variable GDP obtained in the hotel and restaurant sectors. In this case the regression model will no longer contain this variable.

Table 5 - Estimation of standard error deviation
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.908(a)	.824	.789	1012.9791

If we will make the determinations for the new regression model, the result will be:

a. Predictors: (Constant), Tourists arrivals in the establishments of touristic reception (total, thousands of tourists), Investments in hotels and restaurants (millions RON current prices)

In this case the coefficient of determination R^2 indicating the percentage of how much of the total variance is explained by independent variables, is

82.4% (*Table 5*) as compared to 84% in the case of the model with 3 variables (*Table 3*).

This means that the abandoned variable caused a rise of R^2 by 1.6% an extremely small growth for an independent variable. The reach of a standard error of the estimation even lower as compared to the initial model is determined: 1012.97 (table no. 5) as opposed to 11019.05 million RON (table no. 3).

The other processing results are:

Table 6 - Regression coefficients
Coefficients (a)

Model	Indicators	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-16570.909	3580.297		-4.628	.001
	Investments in hotels and restaurants (million RON in current prices)	-6.887	1.669	-1.233	-4.127	.002
	tourists arrivals in the establishments of touristic reception (total, thousand tourists)	5.147	.841	1.829	6.123	.000

a. Dependent Variable dependent: The Gross Domestic Product obtained in the hotel and restaurants sector (millions RON, current prices)

A new regression equation results from the above:

$$\hat{y} = -16570.909 - 6.887\, x_1 + 5.147\, x_2 \quad (4)$$

Table 7 - Analysis of variance
ANOVA (b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	48102936.356	2	24051468.178	23.439	.000(a)
	Residual	10261266.073	10	1026126.607		
	Total	58364202.429	12			

Predictors: (Constant), Investments in hotels and restaurants (millions RON, current prices), Tourists arrivals in the establishments of touristic reception (total, thousands of tourists)

b. Dependent variable: The Gross Domestic Product obtained in the hotels and restaurants sector (millions RON, current prices)

On the basis of the data above a higher value of calculated F can be also ascertained of 23.439 (*Table 7*) as compared to 15.734 (*Table 4*) that enforces the statistical significance of the explanation by regression of variable dependents.

2.1. The issue of multicollinearity

The most wide spread issue that comes up in the case of multiple regression is related to the results of the powerful interdependent connections between independent variables (Lefter, 2004, p.370).

In order to construe correctly the relations between variables one must know the correlation coefficients, which are presented in the following matrix:

Table 8 - Methods of multivariate analysis of dependencies
Correlations

		Gross Domestic Product obtained in the hotel and restaurants sector (millions RON, current prices)	Investments in hotels and restaurants (millions RON, current prices)	Tourists arrivals in the establishments of touristic reception (total, thousands of tourists)	Overnight stays the establishments of touristic reception (total, thousands)
Gross Domestic Product obtained in the hotel and restaurants sector (millions RON, current prices)	Pearson Correlation	1	.406	.724(**)	.757(**)
	Sig. (2-tailed)		.168	.005	.003
	N	13	13	13	13
Investments in hotels and restaurants (millions RON, current prices)	Pearson Correlation	.406	1	.896(**)	.660(*)
	Sig. (2-tailed)	.168		.000	.014
	N	13	13	13	13
Tourists arrivals in the establishments of touristic reception (total, thousands of tourists)	Pearson Correlation	.724(**)	.896(**)	1	.883(**)
	Sig. (2-tailed)	.005	.000		.000
	N	13	13	13	13
Overnight stays the establishments of touristic reception (total, thousands)	Pearson Correlation	.757(**)	.660(*)	.883(**)	1
	Sig. (2-tailed)	.003	.014	.000	
	N	13	13	13	13

The data above shows the existence of certain correlations between independent variables of the model, namely the existence of multicollinearity. This phenomenon distortions the standard error of the estimation and can lead to incorrect conclusions regarding to which of the variable is statistically significant or not.

• The most powerful correlation exists between the investments in hotels and restaurants and tourists arrivals in the establishments of touristic reception since it has the value of 0.896 then between tourists arrivals in the establishments of touristic reception and the, overnight stays the establishments of touristic reception that is 0.883 and finally the correlation between the overnight stays the establishments of touristic reception and investments in hotels and restaurants which is 0.660 (*Table 8*).

It is considered that there are no special problems if the correlation coefficients between the independent variables are within -0.70 and +0.70. Usually, when between the two independent variables the correlation coefficients are high, one of the variables is eliminated from the model (Lefter, 2004, p. 379). The correlation coefficient between the investments in hotels and restaurants and tourists arrivals in the establishments of touristic reception is 0.896. Also, the correlation coefficient between the tourists arrivals in the establishments of touristic reception and overnight stays the establishments of touristic reception is 0.883 and it is significant at a level of significance of 0.05 (*Table 8*).

In order to avoid any distortion phenomenon the variable tourists overnight stays in the establishments of touristic reception can be given up, for example.

According to the results obtained by multiple regression analysis both investments and arrivals of tourists are an important predictor of GDP.

It is a paradox that Romania failed to make tourism a key sector despite the fact that it has a complex and varied tourist potential, having the "raw material" to develop particular tourist products such as: mountain and cultural circuits, spa and rural tourism etc. In order to know the more demanding requirements of tourists' quantitative and qualitative marketing research should be made periodically. Qualitative research among managers of tourist facilities and quantitative research among tourists for a profound knowledge of the consumers of travel services, all of which require investments, but I am firmly convinced that such investments would have a positive contribution to the evolution of GDP.

The companies in Romanian tourism should emphasize more on marketing policies that can be also found at the level of marketing strategies. In my opinion it should be understood that the differentiation strategy through staff involved in the provision and marketing of tourism services plays a crucial role in attracting and maintaining customers. I also believe that considering the conditions of an increasingly powerful competition, the entrepreneurs in Romanian tourism should use continuous renewal of their offer, because this way they can adapt products and services to the increasingly demanding requirements of tourists.

3. CONCLUSION

As a consequence, it can be ascertained that the variables:

• Investments in hotels and restaurants;

• Tourists arrivals in the establishments of touristic reception are significant predictors for the GDP dependent variable obtained in the hotel and restaurant sector, in the analyzed period 1995-2007.

The most powerful correlation is between:

• The GDP obtained in the hotels and restaurants sector and, overnight stays the establishments of touristic reception of 0.757,

• The GDP obtained in hotels and restaurants sector and tourists arrivals in the establishments of touristic reception 0.724 and finally,

• The GDP obtained in the hotels and restaurants sectors and investments in hotels and restaurants, of 0.406.

Among independent variables the most powerful correlation occurs between investments in hotels and restaurants and tourists arrivals in the establishments of touristic reception, this being 0.896.

The number of employees in tourism has increased but unfortunately without their training. The managers of tourist facilities have not been motivated to change their staff or to invest in training sessions. In order to resist competition in the next years, I believe that the focus should be on investments in hotels and restaurants, this variable having a strong correlation with the variable of tourists' arrivals in the establishments of touristic reception.

Inevitably the following question arises: what are those components of tourism marketing mix that would help to increase the number of tourists arrivals in the establishments of touristic reception?

I do not deny that promotion is not a particularly important element in attracting customers and certainly the number of tourists might increase due to the promotion of Romania in the cold season in underground stations in Rome and Milan, action that has cost nearly 100.000 Euros and this is not the only international campaign to promote tourism to our country - in Paris, the Ministry of Tourism has launched and Romania land of choice campaign - but my personal belief is that more importance should be given to investments in expanded marketing mix components, more precisely to investments in people, processes and in physical evidence that would help retain tourists.

Even if the posters posted in Milan and Rome will have a positive effect and they will contribute to increasing tourist arrivals during the winter if we do not pay increased attention and we do not invest in the three elements of the specific marketing mix and not only, there is a risk to "lose" these tourists won by promotion.

Strategies for diversification, differentiation, renewal of products, of flexibility should have an increased application in future.

BIBLIOGRAPHY

1. Lefter, C. (2004) *Marketing researches,* Published by Infomarket, Braşov.
2. Popescu, C. (Grant Manager), Neguţ, S., Bran, F., Alexandru, F., Ioan, I., Rădulescu, C.V., Dogaru, D, Dragne, D.M., Necşuliu, (Borto) G., Simion, G.*,* (2005) *The role of touristic activity in regional development of Romania ,* Journal of Science Politics and Scientometry - Special Edition, p. 8.
3. *** (2008) *Romanian Statistic Year Book*, pp. 477, 517, 826-827.
4. *** (2008) *The system of statistic indicators of tourist activity*, http://www.biblioteca-digitala.ase.ro/biblioteca/pagina2.asp?id=cap3.
5. *** http://www.insse.ro/cms/rw/pages/index.ro.do, 2009.