Analysis of Main Economic Factors Influence on Romanian Tourists Number Accommodated in Romania, using Anova Method

Emilia Gabroveanu
Romanian-American University, București
gabroveanu.emilia@profesor.rau.ro
Roxana Elena Stan
Romanian-American University, București
stan.roxana.elena@profesor.rau.ro
Nicoleta Radneantu
Romanian-American University, București
radneantu.nicoleta@profesor.rau.ro

Abstract. *Now, it is indisputable the role, increasingly important, that tourism plays in economic and social development. This is reflected particularly through the tourism movement, which by generating revenue contributes to the creation of GDP.*

Size of tourist movement can be expressed through the following indicators: number of tourists, the number of check-in for nights, average length of holiday, density of tourist movements, the relative preference for travel, the revenue and the average number of tourists.

The aim of this paper is to identify the main influence of economic factors – total income of households and consumer price indices, registered in Romania in the period 2001-2007 –, on the evolution Romanian tourists number accommodated in Romania, using the dispersion analysis method ANOVA.

Keywords: Romanian tourists; total income of households; consumer price indices; multifactorial regression model; variation analysis.

JEL Codes: C19, M00.
REL Codes: 7Z, 10B, 14F.

1. Introduction

Considered social-economic phenomenon specific to modern civilization, in recent decades, tourism has become one of the most important economic activities.

As demand for products in general, demand for services depends on a number of factors, and is derived from the consumer behavior which, under the restrictions of its budget, wants to maximize utility.

In literature, the experts grouped the factors that influence tourist's movement in leisure, income, offer and motivation. In practice however, the number of these factors is greater. The main economic factors which influence the demand for tourist products, respectively tourist movement are income and prices/tariffs.

Considered the main condition of tourist's movement existence, revenues are influencing the size and frequency with which it is manifested tourism demand. Hicks, in Value and Capital paper, introduce the concept of dynamic economic equilibrium. It emphasizes that any action is changed continuously by its effects, changing conditions generating the modification of equilibrium.

Prices and tariffs are another factor stimulating tourism. Prices and tourism demand are interrelated, due to tourism movement instability, the rigidity of supply, different manifestation of segments of customers.

Commensurable methods of the phenomena make possible the concrete assesses of hypostasis registered at micro and macroeconomic levels. It is necessary to describe each examined situation, the comparison between two or more collectivities and analysis the influence of involved factors.

A method of dispersion analysis is the ANOVA method – Analysis of Variance – contoured by A. Fisher, involving grouping and studying interdependent that occurs during the event of economic phenomena and processes. Study of these connections involves building a regression model.

2. Defining econometric model

Due to permanent changes taking place at both macro and microeconomic, is necessary description and analysis of economic and social phenomena and processes and analyze the influence of factors which, in time, may change them. To highlight the degree of influence of different factors may apply multifactorial regression model, respectively dispersion analysis method *ANOVA – Analysis of Variance*.

ANOVA method implies defining a regression model, which involves: determining the regression equation, testing the validity of the regression model, establishment and testing correlation, and testing significance and estimation of linear regression model parameters on confidence intervals.

The *identification of regression equation* which shapes the link between the analyzed characteristics implies:

- identify variables: x_i and y_i

x_i – factorial variables;

y_i – resultant variables.

- estimate parameters of linear regression model.

Linear regression model, where the analysis to be done, is based on linear regression equation, which is:

$$y_i = b_0 + b_1 \times x_{i1} + b_2 \times x_{i2}. \quad (1)$$

Parameters are determined, as with unifactorial model of linear regression, applying the least squares method. From the first order conditions, it is obtained the linear equations system:

$$\begin{cases} n \times b_o + b_1 \times \sum_{i=1}^{n} x_{i1} + b_2 \times \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i \\ b_o \times \sum_{i=1}^{n} x_{i1} + b_1 \times \sum_{i=1}^{n} x_{i1}^2 + b_2 \times \sum_{i=1}^{n} (x_{i1} \times x_{i2}) = \sum_{i=1}^{n} (y_i \times x_{i1}) \\ b_o \times \sum_{i=1}^{n} x_{i2} + b_1 \times \sum_{i=1}^{n} (x_{i1} \times x_{i2}) + b_2 \times \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} (y_i \times x_{i2}) \end{cases} \quad (2)$$

Testing the validity of the regression model can be achieved with the ANOVA dispersion analysis method. This implies going through the following steps:

- establishing assumptions:

- Null hypothesis – H_0: the model is not valid

- Alternative hypothesis – H_1: the model is valid.

- applies F test (Fisher Snedecor) in relation to data;

- is set the significance á limit and it is obtained the rejection limit;

- determine the statistics of F test;

- it takes final decision in testing the validity of the model depending on the Fc value:

- if $F_c > F_{\alpha;k;n-k-1}$, then H_0 is rejected, the regression model is statistically valid;

- if $F_c < F_{\alpha;k;n-k-1}$, then H_0 is considered true, the regression model is not statistically valid.

In statistics, the *multiple correlations* are used to measure the intensity of the connection between a factorial variable and resultant variable. In econometrics the multiple correlations is introduced also in the phase of linear regression model validation. If the result is closer to 0, the connection is weak (if R = 0, between variables is no connection), and a value closer to 1 indicates a stronger connection (if R = 1 there is a perfect connection between variables).

Testing significance of linear regression model parameters on confidence intervals has in view the collectivity size, respectively the statistical units observed. Since n < 30, it is applied Student test, and if n > 30 one applies Z test.

3. Building the regression model

To determine and analyze the degree of dependency between the number of Romanian tourists accommodated in Romania, and total income of households and consumer price indices, it will be applied multiple regression method.

During 2001-2007, in Romania, the analyzed characteristics have registered the following evolution:

Evolution of Romanian tourist's number, total income of households, and consumer price indices registered in Romania

Table 1

Years	Romanian tourists number * (thousands tourists)	Total income of households (lei, monthly per household)	Consumer price indices **
2001	3960	521.79	100
2002	3848	658.51	126.1
2003	3952	795.09	119.3
2004	4280	1085.79	112.6
2005	4375	1212.18	112.7
2006	4836	1386.32	105.6
2007	5421	1686.74	102.89

* Romanian tourist's arrivals in the establishments of touristic reception with functions of touristic accommodation in Romania.

** Paid accommodation at hotel.

Source: Romanian Statistical Yearbook, National Institute of Statistics, 2008, pp. 826, 460.

Here are the steps followed for dispersion analysis method ANOVA. To facilitate the calculations and problem solving it has been used Excel software.

Determine the regression equation

Analyzing characteristics: Romanian Tourist's number, total income of households and consumer price indices (see Table 1), determine:

y_i –the number of Romanian tourists (arrivals of Romanian tourists in the establishments of touristic reception with functions of touristic accommodation - thousands tourists)

x_{i1} – total income of households (lei, monthly per household)

x_{i2} – consumer price indices (paid accommodation at hotel)

Regression equation for model the link between the variables mentioned is:

$$y_i = b_0 + b_1 \times x_{i1} + b_2 \times x_{i2}.$$

In Table 2, are presented intermediate values necessary for apply the least squares method:

Algorithm required for applying least squares method

Table 2

Years	x_{i1}^2	x_{i2}^2	$x_{i1} \times x_{i2}$	$y_i \times x_{i1}$	$y_i \times x_{i2}$
2001	272264.80	10000	52179	2066288.40	396000
2002	433635.42	15901.21	83038.11	2533946.48	485232.80
2003	632168.11	14232.49	94854.24	3142195.68	471473.60
2004	1178939.92	12678.76	122259.95	4647181.20	481928
2005	1469380.35	12701.29	136612.69	5303287.50	493062.50
2006	1921883.14	11151.36	146395.39	6704243.52	510681.60
2007	2845091.83	10586.35	173548.68	9143817.54	557766.69
$\sum$	8753363.57	87251.46	808888.06	33540960.32	3396145.19

Knowing that n = 7, $\Sigma x_{i1} = 7346.42$, $\Sigma x_{i2} = 779.19$ and $\Sigma y_i = 30672$, the system equations become:

$$\begin{cases} 7 \times b_0 + 7346.42 \times b_1 + 779.19 \times b_2 = 30672 \\ 7346.42 \times b_0 + 8753363.57 \times b_1 + 808888.06 \times b_2 = 33540960.32 \\ 779.19 \times b_0 + 808888.06 \times b_1 + 87251.46 \times b_2 = 3396145.19 \end{cases}$$

Where: $b_0 = 4807.61$, $b_1 = 1.17$ and $b_2 = -14.85$.

Regression line is defined by the next equation: $y_i = 4807.61 + 1.17 \times x_{i1} - 14.85 \times x_{i2}$.

Testing the validity of the regression model

To confirm the establishment and use the model the approach was to verify the validity of the method. This is achieved in the following stages of analysis of linear regression model.

The established assumptions are:

- H_0: the model is not valid (null hypothesis)
- H_1: the model is valid (alternative hypothesis).

To use the F test, compared with the data was applied EXCEL software. The values obtained, necessary for analysis of the validity of regression model, are:

ANOVA table with the necessary value for analysis the regression model

Table 3

ANOVA					
	Df	SS	MS	F	Significance F
Regression	2	1846919	923459,4	37.95923	0.002505
Residual	4	97310,66	24327,66		
Total	6	1944229			

Knowing that the probability of occurrence of the results is 95% and the limit of significance $\alpha = 0.05$ it was obtained the limit region of rejection:

$$F_{\alpha;k;n-k-1} = F_{0.05;2;4} = 6.94$$

For examined characteristics (Romanian tourists number, total income of households, consumer price indices), $F_c = 37.95923 > F_{\alpha;k;n-k-1} = 6.94$. In this case H_0 is rejected; the regression model is statistically valid.

One other element which reflects the validity of the model is Significance F. If the value get for this element is less than 0,05, $H_0$is rejected. In Table 3 it is shown that Significance F = 0.002505 < 0.05, and the regression model is statistically valid.

Determining and testing the correlation

To determine correlation Excel software was used. The values obtained for the correlation (R), R Square (R^2), the adjusted R Square (R^2) and standard deviation of errors (s_e) are presented in Table 4:

The values of parameters R, R^2, $\overline{R}^2$ and s_e

Table 4

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.974653
R Square	0.949949
Adjusted R Square	0.924923
Standard Error	155,9733
Observations	7

The correlation value (Multiple R) is set to 0.974653, and shows that between the number of Romanian tourists and total income of households, and consumer price indices registered in Romania in the period under review, there is a very strong connection.

R Square is set to 0.9499 and shows that the influence of the two factors (total income of households and consumer price indices) on the number of Romanian tourists accommodated in Romania is 94.99%, the remaining 5.01% is due to other random factors.

The value of standard error is 3165.991, and it shows that points are not placed near the regression right.

Testing the significance and estimation of parameters for linear regression model on confidence intervals

Estimators for regression parameters are determined by least squares method, and testing their significance has in view the size of the selected collectivity. Since n <30, it is applied Student test.

In Table 5, an ANOVA table, are given the necessary elements for testing parameters and establish confidence intervals.

Elements necessary for testing parameters and establish confidence intervals

Table 5

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	4807.605	907.8452	5.295622	0.006105	2287.023	7328.188
X Variable 1	1.168771	0.165181	7.075682	0.002105	0.710154	1.627387
X Variable 2	-14.8456	7.416283	-2.00175	0.115884	-35.4365	5.745336

The above table shows the linear adjusting equation of the connection between indicators:

$$y_i = 4807.61 + 1.17 \times x_{i1} - 14.85 \times x_{i2}$$

The value of b_0 is 4807.61 shows that if total income of households is remaining constant, respectively consumer price indices, the number of Romanian tourists accommodated in Romania was set to a value of approximately 4808.

Analyzing the result for t test, $t_c = 5.295622$, and significance threshold P-value = 0.006105 < 0.05, we see that free time is statistically significant. This is reflected by the presence of the confidence interval, respectively $2287.023 < b_0 < 7328.188$.

b_1 coefficient value is approximately 1.1688 and shows that by increasing of 1 leu monthly per household, the number of Romanian tourists will increase by approximately 1 tourist. For this coefficient threshold of significance P-value is 0.002105 < 0.05, respectively $t_{cb1} = 7.075682 > 2.571 = t_{0.025;5}$, the coefficient is considered statistically significant.

The lower limit of the confidence interval of this parameter is approximately 0.710154, and the upper limit reaches 1.627387.

The coefficient b_2 is -14.8456 and it shows that the decrease of 1% of consumer price indices will result in the decrease of the number of Romanian tourists with approximately 15 tourists. The threshold of significance P-value is 0.115884> 0.05, and the result $t_{cb2} = -2.00175 < 2.571 = t_{0.025;5}$ shows that this coefficient is not statistically significant.

4. Conclusions

Analysis carried out on recent developments that took place in the tourism sector reflects the trend of increasing the number of domestic consumers of tourism products.

After testing the validity of the regression model, it was established that this model is statistically valid. Thus, it was observed that between the number of Romanian tourists and total income of households, and consumer price indices registered in Romania in the period under review, there is a very strong connection, correlation has the value 0.974653.

The influence percentage of total income of households and consumer price indices, reflected by R Square, on the number of Romanian tourists accommodated in Romania is approximately 94.99%, the remaining 5.01% being the result of other random factors.

After the analysis performed it was established that the increase of 1 leu monthly per household will cause an increase of the number of Romanian tourists by approximately 1 tourist and the decrease of 1% of consumer price indices will result in the decrease of the number of Romanian tourists with approximately 15 tourists.

Generally, lower prices lead to increased tourism demand (increased tourists number, length of stay, frequency of travel etc.), and vice versa. However practice has shown that price is not the main motivator.

Holders of accommodation units in Romania have seen a growing number of customers, most of them being Romanian tourists. To Romanian tourists are added as a barrier for transit are removed and is created the necessary infrastructure, ever more foreign tourists.

References

Andrei, T. (2003). *Statistică și econometrie*, Editura Economică, București

Chow, G. (1989). *Econometrics*, McGraw Hill, New York

Dougherty, Ch. (2007). *Introduction to Econometrics*, Oxford University Press

Gogonea, R., Zaharia, M. (2008). *Econometrie cu aplicații în comerț-turism-servicii*, Editura Universitară, București

Hannan, E.J., Deistler, M. (1988). *The Statistical Theory of Linear Systems*, New York Wiley

Pecican, E.Ș. (2007). *Econometrie pentru... economiști*, Editura Economică, București

Reinsel, G.C. (1997). *Elements of Multivariate Time Series Analysis*, Springer Varleg, New York

Wonnacott, T.H., Wonnacott, R.J. (1990). *Introductory Statistics for Business and Economics*, New York Wiley

Romanian Statistical Yearbook, National Institute of Statistics, 2008