

doi: 10.52370/TISC26450BC

REGRESSION ANALYSIS IN EXAMINING FACTORS AFFECTING TOURISM REVENUES

Bojana Ćendić¹; Zorica Sredojević²; Nataša Milosavljević³

Abstract

In this paper, the application of multiple linear regression is analysed to assess the factors affecting tourism revenues in European cities. The study uses hypothetical yet realistic data for 15 cities in 2019. Average hotel price, number of tourists, and number of international air connections are included as independent variables, while total tourism revenue is considered as the dependent variable. The regression model was developed and tested using SPSS, estimating coefficients, standard errors, and t-values. Multicollinearity was assessed using the Variance Inflation Factor (VIF), and residual analysis was conducted to confirm model assumptions and prediction reliability. Results show that the number of tourists has the greatest impact on revenue, while other variables represent complementary indicators. The model is statistically significant ($p < 0.001$) and explains 99.6% of revenue variability, confirming its high predictive accuracy and practical relevance for planning tourism development.

Key Words: *multiple regressions, tourism, revenue, SPSS, statistical analysis, prediction*

JEL classification: *C20, C51, L83*

¹ Bojana Ćendić, PhD, Assistant Professor, State University of Novi Pazar – Department of Natural and Mathematical Sciences, Vuka Karadžića bb, 36300 Novi Pazar, Serbia; PhD, Professor of Professional Studies, Belgrade Business and Arts Academy of Applied Studies, Kraljice Marije 73, Zvezdara, 11000 Belgrade, Serbia, +381 65 221 3142, bojana.cendic@gmail.com

² Zorica Sredojević, PhD, Full Professor, University of Belgrade - Faculty of Agriculture, Nemanjina 6, 11080 Belgrade - Zemun, Serbia, +381 69 4457 412, zokas@agrif.bg.ac.rs

³ Nataša Milosavljević, PhD, Associate Professor, University of Belgrade - Faculty of Agriculture, Nemanjina 6, 11080 Belgrade - Zemun, Serbia, +381 64 127 1879, natasam@agrif.bg.ac.rs

Introduction

Tourism is currently one of the most important sectors of the global economy, exerting a significant influence on economic development, employment, and international exchange. Its expansion contributes to economic growth and plays a significant role in regional development (Zhang & Wen, 2022). The development of the tourism sector depends on a wide range of factors, among which the number of tourists, the price level of tourism services, and the availability and quality of infrastructure are particularly prominent. Yan & Zhou (2024) examine the factors influencing tourism revenues using multiple linear regression, including accommodation prices, the number of hotels, and the number of tourists. Montes-Rojas & Barroso (2020) show that increased air connectivity is associated with higher tourist arrivals. In contemporary tourism research, quantitative methods-particularly regression-based approaches-have assumed an increasingly important role, enabling a more precise analysis of the relationships among relevant variables. Previous studies widely apply such approaches to examine economic relationships within the tourism sector (Wen, 2025; Chen et al., 2023). In addition to individual determinants, increasing attention is being given to the broader context in which tourism develops.

Purwono et al. (2024) investigate the relationship between destination competitiveness and tourism performance, highlighting the importance of infrastructure, connectivity, and core resources. Furthermore, Mohebi & Rahim (2010) find the determinants of tourism expenditure, showing that income, prices, and travel costs influence tourism expenditure. Although a substantial body of recent research has focused on the effects of the COVID-19 pandemic on tourism demand and sector performance (Kumar & Ekka, 2024), this study focuses on the pre-pandemic period, specifically the year 2019, which can be considered representative for analysing tourism in the absence of major global disruptions. This approach enables a clearer examination of the structural economic relationships among the selected variables. Tourism revenues represent one of the key indicators of tourism sector performance. Yan & Zhou (2024) apply multiple linear regression to analyse the factors influencing tourism revenues, including accommodation prices, the number of hotels, and the number of tourists. A similar approach is adopted by Zhang & Wen (2022), who examine the tourism sector using factor and regression analysis. Wen (2025) further analyses the role of economic and infrastructural factors in tourism development. Chen et al. (2023) focus

on the impact of various factors on tourism revenues using uncertain regression analysis. Montes-Rojas & Barroso (2020) analyse the determinants of international tourist arrivals and expenditure, showing that gross domestic product, air connectivity, and relative prices are associated with tourism outcomes. Mohebi & Rahim (2010) explain that income has a positive effect on tourism expenditure while prices and travel costs have a negative influence. Li et al. (2024) analyse differences in tourism economic development and their determinants using multiple linear regression and GIS-based spatial analysis. Kumar & Ekka (2024) assess the impact of the COVID-19 pandemic on tourism demand and mobility, highlighting significant disruptions in tourism activity. The subject of this paper is the analysis of factors influencing tourism revenues in European cities. The independent variables considered include the average accommodation price, the number of tourists, and the number of international air connections, while tourism revenues constitute the dependent variable. The aim of the research is to determine, through the application of multiple linear regression, the relative influence of these factors on tourism revenues.

Materials and Methods

The study is based on hypothetical yet realistic data for a sample of 15 European cities. The selected cities represent different levels of tourism development, thereby ensuring sufficient variability in the data required for the application of regression analysis. For each city, data were collected on the average accommodation price, the number of tourists, the number of international air connections, and total tourism revenues. A quantitative methodology based on multiple linear regression was employed. In order to examine the economic and structural factors influencing tourism revenues, the following variables were defined: tourism revenue (expressed in billions of euros) as the dependent variable, while the independent variables include accommodation price (an indicator of the destination's price level), the number of tourists (an indicator of the volume of tourism demand), and the number of air connections (an indicator of accessibility and transport infrastructure). To analyse the relationships between the variables, a multiple linear regression model of the following form was defined:

$$\text{Revenue} = \beta_0 + \beta_1 \cdot \text{Tourists} + \beta_2 \cdot \text{Price} + \beta_3 \cdot \text{Air connections} + \varepsilon \quad (1)$$

Based on the analysis, the following regression equation was obtained:

$$\text{Revenue} = -0.566 + 1.182 \cdot \text{Tourists} + 0.011 \cdot \text{Air connect.} - 0.012 \cdot \text{Price} \quad (2)$$

The coefficients of the model represent estimates of the effect of each independent variable on revenue, assuming that all other factors remain constant. The statistical analysis was carried out through several interrelated steps. In the first step, basic descriptive statistical measures were calculated in order to examine the structure and variability of the data. Subsequently, the Pearson correlation coefficient was applied to determine the strength and direction of the relationships between variables, as well as for the preliminary identification of potential multicollinearity.

Due to the extremely high correlations observed among the variables, an additional verification of the hypothetical dataset was conducted in order to confirm the consistency of the generated values and exclude possible data-entry inconsistencies. The verification confirmed that the dataset was correctly constructed and that no technical errors occurred during data preparation. The strong correlations can be explained by the fact that major European tourist destinations with a high number of tourists typically also exhibit higher accommodation prices, stronger international air connectivity, and consequently higher tourism revenues. The data were processed using the IBM SPSS Statistics software package, which enables reliable model estimation, statistical significance testing, and verifying the fundamental assumptions of linear regression. The regression model was estimated using the method of Ordinary Least Squares (OLS), yielding the corresponding coefficients, as well as t-values and p-values for testing their statistical significance. The overall significance of the model was assessed using Analysis of Variance (ANOVA), while its reliability was further examined through tests of multicollinearity (Variance Inflation Factor and Condition Index), normality of residuals, homoscedasticity, and the presence of outliers.

Results and Discussion

Data collected from relevant sources on the average accommodation price, the number of tourists, the number of international air connections, and total tourism revenues for the analysed cities are presented in Table 1.

Table 1: *Dataset of analysed European cities*

City	Price (€)	Tourists (millions)	Air connections	Revenue (billion €)
London	150.00	8.00	200	9.30
Paris	145.00	7.20	170	8.10
Rome	140.00	7.50	180	8.50
Barcelona	130.00	6.80	160	7.90
Amsterdam	125.00	6.50	155	7.50
Prague	120.00	6.20	150	7.10
Berlin	115.00	5.80	140	6.70
Vienna	110.00	5.40	130	6.20
Madrid	105.00	5.00	125	5.60
Lisbon	100.00	4.90	120	5.30
Dubrovnik	95.00	4.50	110	4.80
Florence	90.00	4.10	100	4.40
Munich	155.00	8.10	210	9.60
Athens	85.00	3.80	95	3.90
Venice	160.00	8.50	220	10.20

Source: *Authors*

Descriptive statistics for the sample of 15 European cities represent the initial step in the data analysis (Table 2).

Table 2: *Descriptive statistics of variables*

Variable	Mean	Std. Deviation	N
Revenue	7.01	1.96	15
Price	121.67	24.40	15
Tourists	6.15	1.52	15
Air connections	151.00	39.24	15

Source: *Authors*

Based on the analysis results, the average tourism revenue amounts to €7.01 billion, the average accommodation price is €121.67, the average number of tourists is 6.15 million, and the average number of international air connections is 151 (Table 2). The observed variability across cities indicates a high degree of heterogeneity, which justifies the further application of regression analysis to examine the relationships between variables. Relatively high standard deviations indicate substantial variation among cities, particularly in revenues and air connections. In the subsequent analysis, the correlation matrix presented in Table 3 provides insight into the relationships among the observed factors and represents an important step prior to specifying the regression

model. The results indicate a strong positive relationship among all variables, with the correlation between the number of tourists and tourism revenues being particularly high ($r = 0.99$; $p < 0.001$).

Table 3: *Pearson correlation matrix*

Variable	Revenue	Price	Tourists	Air connections
Revenue	1.000	0.992	0.998	0.992
Price	0.992	1.000	0.995	0.989
Tourists	0.998	0.995	1.000	0.991
Air connections	0.992	0.989	0.991	1.000

Source: *Authors*

A high degree of correlation among the independent variables (above 0.98) indicates strong interdependence, representing a clear sign of multicollinearity in the proposed regression model. The results of the regression model estimation are presented in Tables 4 and 5.

Table 4: *Model summary*

Model	R	R Square	Adjusted R Square	Std. Error of the Estim.
1	0.998	0.996	0.995	0.134

Source: *Authors*

The coefficient of determination ($R^2 = 0.996$) indicates that the model explains as much as 99.6% of the variability in tourism revenues.

Table 5: *ANOVA results*

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	53.410	3	17.803	984.623	< 0.001
Residual	0.199	11	0.018	-	-
Total	53.609	14	-	-	-

Source: *Authors*

The adjusted coefficient of determination (Adjusted $R^2 = 0.996$) confirms that the model retains high explanatory power after adjusting for the number of variables. A low standard error of the estimate (0.134) further indicates high model precision and minimal deviations between observed and predicted values. The ANOVA results indicate that the model as a whole is statistically significant ($p < 0.001$). The extremely high F-statistic ($F = 984.62$) confirms that the combination of independent variables has a strong effect on tourism revenue. The analysis of individual variable effects (Table 6) shows that the number of tourists is

the only statistically significant predictor of tourism revenue. The standardised beta coefficient ($\beta = 0.917$) indicates the dominant influence of this variable in the model.

Although the regression model demonstrated very high explanatory power ($R^2 = 0.996$), the Pearson correlation matrix and VIF indicators revealed the presence of multicollinearity among the independent variables. This phenomenon is relatively common in tourism-related economic datasets because tourism demand, pricing levels, and transport accessibility are naturally interconnected. Consequently, the results should primarily be interpreted in terms of the overall predictive performance of the model, while caution is recommended when evaluating the isolated statistical significance of individual predictors.

Table 6: *Regression coefficients*

Variable	B	Std. Error	Beta	t	Sig.	VIF
(Constant)	-0.566	0.388	-	-1.460	0.172	-
Price	-0.012	0.015	-0.144	-0.776	0.454	101.959
Tourists	1.182	0.270	0.917	4.373	0.001	130.227
Air connections	0.011	0.007	0.226	1.619	0.134	57.983

Source: *Authors*

On the other hand, although accommodation prices and the number of air connections are highly correlated with revenue, their individual effects within the model are not statistically significant, which may be explained by pronounced multicollinearity ($VIF > 10$; Condition Index = 136.55). The residual analysis presented in Table 7 indicates that deviations between observed and predicted values are minimal.

Table 7: *Residual statistics*

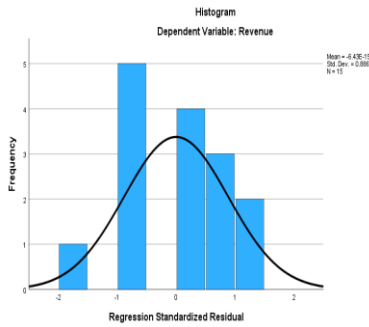
Variable	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4.016	10.117	7.007	1.953	15
Residual	-0.214	0.186	0.000	0.119	15
Std. Predicted Value	-1.531	1.592	0.000	1.000	15
Std. Residual	-1.591	1.384	0.000	0.886	15

Source: *Authors*

Residuals range from -0.214 to 0.186 billion Euros, while standardised residuals do not exceed the critical thresholds of ± 3 . This confirms the absence of outliers and the stability of the model. Visual analysis of the fulfilment of all main regression conditions was conducted using a

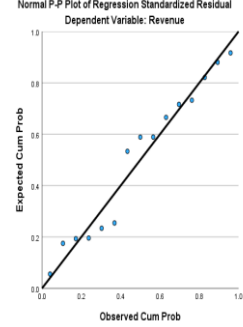
histogram (Figure 1), which indicates a normal distribution of residuals, a P–P plot (Figure 2), agreement with the theoretical distribution, and a scatter plot (Figure 3), which confirms homoscedasticity, thus further supporting the reliability and precision of the proposed regression model.

Figure 1: *Histogram*



Source: *Authors*

Figure 2: *Normal P-P Plot*



Source: *Authors*

Figure 3: *Scatter plot*



Source: *Authors*

The results of the analysis indicate a very high level of explanation of tourism revenues ($R^2 = 0.996$), confirming that the selected variables have strong predictive power. The number of tourists is the only statistically significant predictor of revenue, confirming its dominant influence relative to the other variables. This finding is consistent with Yan & Zhou (2024), whose results show that the number of tourists has the strongest positive effect on tourism revenues, while Zhang & Wen (2022) identify several factors, including infrastructure and transport, as influencing tourism revenues. Similarly, Wen (2025) analyses the importance of tourism demand and infrastructure, while Chen et al. (2023) focus on specific economic and infrastructural determinants of tourism revenue, although in this study the remaining variables did not show statistical significance. The importance of air connectivity and economic factors is

highlighted by Montes-Rojas & Barroso (2020), while Liu et al. (2022) show that tourism mobility and transportation infrastructure positively influence tourism growth. The findings are also consistent with Mohebi & Rahim (2010), who show that higher prices have a negative effect on tourism expenditure. The results confirm that tourism revenues depend on a broader system of interrelated factors. Li et al. (2024) & Purwono et al. (2024) highlight the importance of tourism resources and infrastructure in explaining tourism performance. The results demonstrate that increases in the number of tourists exert a substantial effect on revenue generation, while other factors function as support within the wider tourism system. This finding is important for tourism development planning, as it implies that strategies should primarily focus on increasing tourist arrivals.

Conclusion

The results confirm that multiple linear regression represent an appropriate and reliable method for analysing the factors influencing tourism revenues. The estimated model demonstrates an exceptionally high level of explained variance ($R^2 = 0.996$), indicating strong predictive capability. The number of tourists is the only statistically significant predictor of tourism revenue, confirming its central importance in shaping the economic effects of tourism. In contrast, accommodation prices and the number of air connections do not exhibit statistically significant individual effects, which may be explained by interdependence among the observed variables and their indirect influence. The results also show that relationships in tourism are not simple but depend on a range of interrelated factors. Although the number of tourists is identified as the dominant factor, the remaining variables should be considered within a broader economic and infrastructural framework. The obtained results suggest that tourism development strategies should focus on increasing tourist arrivals, alongside improvements in infrastructure and destination accessibility. Future research may incorporate real and more extensive datasets, as well as additional variables, in order to enable a more precise analysis of tourism revenues.

References

1. Chen, S., Ning, Y., Wang, L., & Wang, S. (2023). Research on the factors influencing tourism revenue of Shandong Province in China based on uncertain regression analysis. *Mathematics*, 11(21), 4490.

2. Kumar, P., & Ekka, P. (2024). Statistical analysis of the impacts of COVID-19 pandemic on the small and large-scale tourism sectors in developing countries. *Environment, Development and Sustainability*, 26(4), 9625–9659.
3. Li, X., Huang, Y., & Wang, Y. (2024). Differences in tourism economic development and its influencing factors among three major city clusters along the middle reaches of the Yangtze River. *Plos one*, 19(5), e0299773.
4. Liu, J., Yue, M., Yu, F., & Tong, Y. (2022). The contribution of tourism mobility to tourism economic growth in China. *PLoS One*, 17(10), e0275605.
5. Mohebi, M., & Rahim, K. A. (2010). Revenue determinants in tourism market. *American Journal of Applied Sciences*, 7(12) 1593–1598.
6. Montes-Rojas, G., & Barroso, R. (2020). What are the empirical determinants of international tourist arrivals and expenditures? An empirical application to the case of São Tomé and Príncipe. *An empirical application to the case of Sao Tome and Principe. World Bank Policy Research Working Paper*, 9189. World Bank.
7. Purwono, R., Esquivias, M. A., Sugiharti, L., & Rojas, O. (2024). Tourism destination performance and competitiveness: The impact on revenues, jobs, the economy, and growth. *Journal of Tourism and Services*, 15(28) 161–187.
8. Wen, L. (2025). Analysis and empirical study on the influencing factors of the tourism economy in Hunan Province. *Economy*, 12(1), 8–20.
9. Yan, Z., & Zhou, H. (2024). Research on Influencing Factors of Domestic Tourism Revenue based on Multiple Linear Regression Analysis. *Highlights in Science Engineering and Technology*, 107, 60–64.
10. Zhang, W., & Wen, L. (2022). An examination of the variables affecting the growth of the tourist sector in Guizhou province. *Sustainability*, 14(18), 11297.