Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development
Vol. 23, Issue 3, 2023
PRINT ISSN 2284-7995, E-ISSN 2285-3952

STATISTICAL ANALYSIS MODEL FOR IDENTIFYING THE DYNAMICS OF HOTEL TOURISM

Elena COFAS[1] , Florin Cristian CIOBĂNICĂ[2]

[1]University of Agronomic Sciences and Veterinary Medicine Bucharest of Bucharest, 59 Marasti Boulevard, District 1, 011464, Bucharest, Romania, E-mail: cofas.elena@managusamv.ro
[2]Theoretical High School "Mihail Kogălniceanu", 8 Narciselor Street, Snagov, Ilfov County, Romania, E-mail: cristiciobanica@lmk.ro

Corresponding author: cofas.elena@managusamv.ro

Abstract

Data analysis is a process of inspecting, transforming, cleaning, and modeling information with the purpose of uncovering useful insights, providing conclusions, and supporting decisions. The results obtained from processing statistical data are presented as specific statistical models (series, graphs, tables), and the relationships for observed phenomena emerge in a sequence that corresponds to the existing relationships. These make it possible to interpret patterns, enabling the correct selection of methodologies for calculating statistical indicators. Statistical data in tourism provide the opportunity for the development of strategies within this sector, where evaluated indicators serve as a source for monitoring and refining situations in tourism. The assessed information allows for analyses across various periods, integrating tourism into the national statistical system. The purpose of this study is to present an analytical model through a phased realization of statistical analysis regarding tourist arrivals in hotel-type accommodation establishments. Essentially, this analysis includes: the statistical study of frequency distributions through the calculation of variation and asymmetry indicators, the statistical study of frequency distributions through the calculation of central tendency indicators, and the establishment of statistical relationships between the two characteristics – the number of hotels and the number of tourists arriving in hotels.

Key words: *hotel tourism, dynamics, statistical analysis model, Romania*

INTRODUCTION

Tourism stands as a pivotal and substantial industry in a country's economy. Romania's wealth of natural, climatic, and historical attributes bestows considerable advantages upon tourism compared to other sectors of the economy. Fostering the growth of the tourism sector should take precedence on both a national level and among entrepreneurs within this domain. It is imperative to promote important decisions and escalate efforts towards developing regions with pronounced tourism potential. Presently, the tourism industry is experiencing a robust upswing following the easing of restrictions post the SARS-CoV-2 pandemic, as the population eagerly seeks to resume their ordinary lives.

The primary objective of statistics is to provide high-quality data and information that embody accuracy, relevance, accessibility, timeliness, coherence, and clarity. Statisticians' provided data offers the capacity to evaluate decision-making factors more broadly. Many accommodation proprietors remain unaware of the opportunities that statistical data can unlock. Statistical data within the tourism realm extend the potential for devising strategies within this sector, with evaluated indicators serving as a resource for monitoring and refining the tourism landscape.

The assessed information enables the execution of analyses across diverse timeframes, effectively integrating the tourism sector into the national statistical framework.

In summary, the significance of statistics in tourism encompasses [2]:

-understanding society as a whole and the level of economic development;

-establishing directions and development objectives;

-formulating prospective and developmental programs;

-determining measures to be taken within the decision-making process;

Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development
Vol. 23, Issue 3, 2023
PRINT ISSN 2284-7995, E-ISSN 2285-3952

-monitoring the achievement of established objectives;
-disseminating the obtained data;
-conducting international comparisons.
In this context, the purpose of this research was to develop an analytical model through a phased realization of statistical analysis regarding tourist arrivals in hotel-type accommodation establishments in Romania.

MATERIALS AND METHODS

A tourist accommodation facility is any permanent or seasonal construction or arrangement that provides accommodation services and other specific amenities for tourists.
In statistical research, tourist accommodation facilities with an accommodation capacity of fewer than 5 places are not included.
A hotel is a type of tourist accommodation facility set up in buildings or building structures that offers properly equipped rooms, studios, or apartments to tourists.
It provides specialized services, has a reception area, and may include dining spaces.
Tourist arrivals in tourist accommodation facilities refer to the number of tourists staying in tourist accommodation establishments and encompass all individuals (both domestic and foreign) travelling outside their permanent place of residence and spending at least one night in a tourist accommodation facility in visited areas within the country.
In practice, this study utilizes two sets of data covering the period from 2005 to 2022 (Table 1):

1. tourist accommodation capacity (the first analyzed characteristic being the *number of hotels*) and
2. the utilization of tourist accommodation facilities, measured by the number of tourist arrivals in hotels (the second analyzed characteristic being the *number of tourists*).

The presented data constitutes an information set concerning the number of hotels (Xi) and the number of tourists/ arrivals in hotels (Yi) on a national level spanning the period from 2005 to 2022. These data (Xi and Yi characteristics) can be subjected to statistical analysis to identify trends and patterns in tourism dynamics within this region.

Table 1. Statistically analyzed characteristics

Year	Number of hotels (Xi)	Number of tourists/ arrivals in hotels (Yi)
2005	989	4,477,936
2006	1,059	4,725,448
2007	1,075	5,212,170
2008	1,104	5,245,292
2009	1,159	4,539,858
2010	1,233	4,585,211
2011	1,308	5,357,763
2012	1,384	5,740,304
2013	1,429	5,908,649
2014	1,456	6,314,865
2015	1,522	7,214,613
2016	1,530	7,927,540
2017	1,577	8,565,979
2018	1,616	9,004,486
2019	1,608	9,274,954
2020	1,581	4,116,681
2021	1,583	6,203,693
2022	1,602	7,943,950

Source: NIS, 2023, http://statistici.insse.ro:8077/tempo-online/[8].

Based on these two datasets, the following analyses can be conducted:

- ✓ statistical analysis of frequency distributions through the calculation of ***variation indicators*** (absolute range of variation, relative range of variation, absolute individual deviations, relative individual deviations, mean absolute deviation, variance, standard deviation, coefficient of variation) and ***skewness indicators*** (absolute skewness, relative skewness, skewness coefficient);
- ✓ statistical analysis of frequency distributions through the calculation of ***central tendency indicators*** (measures of central tendency - arithmetic mean, harmonic mean, geometric mean, quadratic mean, as well as positional average indicators - median and mode);
- ✓ establishing statistical relationships between variables (the two characteristics), both through simple methods (graphical method, correlation table method, grouping method, method of interdependent parallel series) and through analytical methods

(regression method, covariance method, correlation ratio method, correlation coefficient method, analysis of variance method):

→ *The grouping method* is a technique for organizing data whereby statistical relationships can be investigated. Simple grouping involves organizing statistical units based on a primary grouping characteristic and calculating and interpreting partial means or partial relative values for the resultant characteristic. Combined grouping involves dividing statistical units into groups based on variations of two grouping characteristics, and the results of the grouping are presented in a two-way combined table.

Depending on the type of grouping variable (discrete or continuous) and the range of characteristic values, grouping can be done by categories (when data is grouped according to a discrete variable with a relatively small range of values) or by intervals (when data is organized according to a continuous variable with a wide range of values). This method should be employed only when dealing with a large number of statistical observations, where applying analytical calculation methods is not feasible without prior data grouping.

→ *The graphical method* involves constructing a correlation graph (known as a scatterplot) where the values of the primary grouping characteristic (x) and the values of the secondary grouping characteristic (y) are plotted. Based on the distribution pattern of points on the graph, a line or curve is visually drawn, the equations of which are known. If the curve or line is drawn on the first diagonal, the relationship is direct, and if it's drawn on the second diagonal, the relationship is inverse.

→ *The regression method* is an approach to analyzing relationships between variables using regression functions. Regression functions are selected empirically using correlation graphs and significance tests. Based on the number of variables used in the model, two types are distinguished: single-factor regression (one predictor variable xi and one response variable yi) and multiple-factor regression (multiple predictor variables and a single response variable).

→ *The correlation coefficient method* revolves around the "correlation coefficient," which serves as a synthetic indicator used to measure relationships between two variables with a normal distribution.

→ *The correlation ratio method* is used to determine the strength of relationships between the considered influencing factor (xi) and the resultant characteristic (yi), regardless of the form of the relationship—whether linear or nonlinear—by calculating a synthetic correlation indicator called the "correlation ratio" (Rx/y). By measuring the degree of intensity, the link between the characteristics considered as influencing factors and the resulting characteristics is established, without concerning whether the relationships are linear or nonlinear.

→ *The analysis of variance method* can be employed in several cases, namely: to verify the independence of a commercial or tourism-related phenomenon, to assess the stability of means and variances across multiple successive samples, and to examine the dependence of a commercial or tourism-related phenomenon on its influencing factors. All statistical processing and calculations were conducted using the statistical analysis software ANOVA and the Excel function library, ensuring easy and accurate results generation.

RESULTS AND DISCUSSIONS

To characterize the relationship between phenomena, several approaches fall under the category of simple methods for characterizing connections. These methods are easy to apply and rely on qualitative analysis of correlated variables, offering insights into the nature and essential traits of the investigated relationship. Utilizing graphical methods, a quick visual analysis of the trend of the two analyzed characteristics can be achieved, as shown in the following two graphs. In Figure 1, the evolution of the total number of major accommodation establishments in Romania is depicted. It is evident that agrotourism guesthouses are leading by a significant margin, followed by hotels and tourist guesthouses. Agrotourism guesthouses and

Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development
Vol. 23, Issue 3, 2023
PRINT ISSN 2284-7995, E-ISSN 2285-3952

hostels have become increasingly popular and appreciated in recent years, with a doubling of accommodation spaces between 2005 and 2022. Hotels and tourist guesthouses have also experienced modest growth, while cabins and vacation villas have not expanded their accommodation capacity. The general trend regarding the number of hotels has been a gradual annual increase, relatively stable over time, with small fluctuations in certain periods.

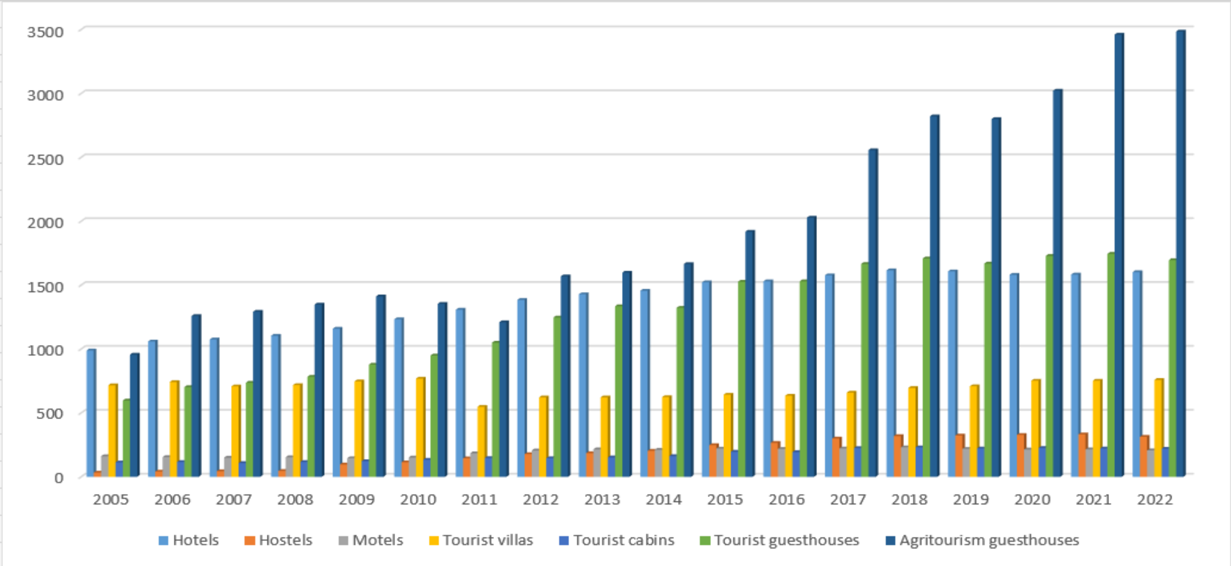

Fig. 1. The total number of accommodation spaces for the main lodging establishments during the period 2005-2021
Source: NIS, 2023, http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table [8].

In Fig. 2, it's evident that the largest number of tourists prefer to stay in hotels, with numbers increasing from over 4.4 million in 2005 to a peak of 9.2 million in 2019. The second preferred accommodation choice is tourist guesthouses, accommodating over 580 thousand people in 2012, with the peak being in 2019 at 1.2 million. There are two notable points of reference when a decline in tourist numbers occurred, namely during the global economic crisis (2008-2010) and during the SARS-Cov-2 pandemic, which resulted in severe travel restrictions (2020-2021). These restrictions started to ease in 2022, allowing for a resumption of travel and vacations, and consequently, a significant increase in the number of tourists (some of the 2022 data might not be updated, which is why this year is missing from the graph representation).

In actuality, the number of tourists experienced a 17% upswing from 2005 to 2008, followed by a remarkable surge of 202% between 2010 and 2019, culminating in a peak in 2019. The onset of the SARS-CoV-2 pandemic and the subsequent travel bans led to a stark plummet of 207% in tourist numbers in just one year. However, commencing from 2021, as travel restrictions eased, there was a noteworthy resurgence, with a 68% upswing in tourist numbers. This trend has remained consistent and is ongoing.

In Figure 3, the two characteristics are depicted in the form of a scatterplot, aiming to observe and potentially establish a connection between them. The relationships that can emerge are stochastic relationships, where one phenomenon acts as an influencing factor and the other as an effect.

Statistical dependence is characterized by the fact that when an influencing factor changes, the influenced factor responds with a distribution of values. As apparent from the parallelism of the two trends, the connection between variables is direct.

However, since there isn't a concrete manifestation of a relationship between the two variables (xi and yi) that can be quantitatively expressed to measure the intensity of their occurrence, a statistical analysis using specific analytical methods is necessary.

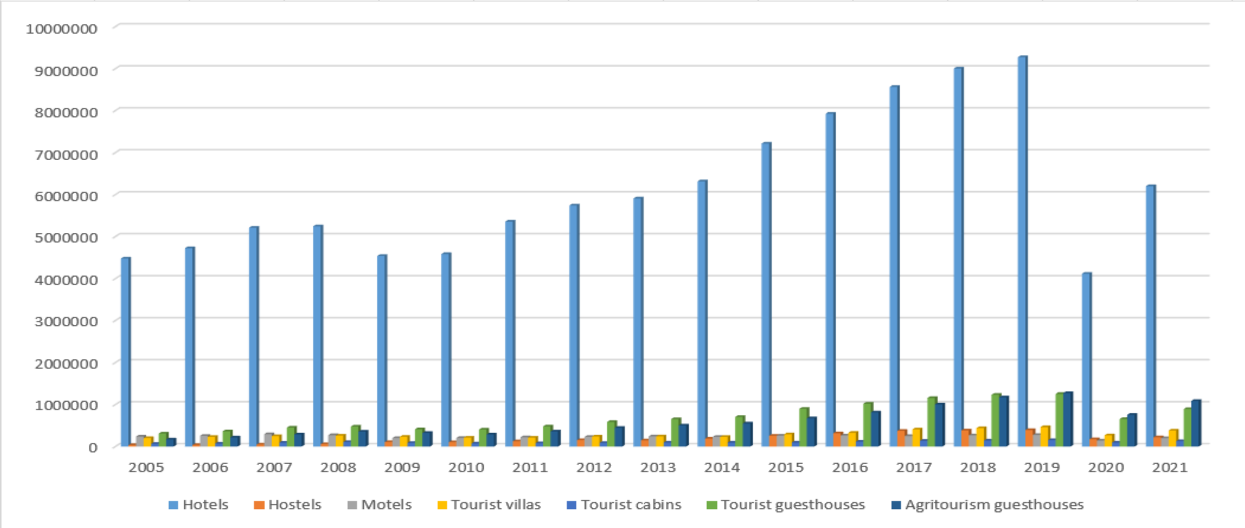

Fig. 2. The total number of tourists during the period 2005-2021, categorized by the main lodging establishments
Source: NIS, 2023, http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table [8].

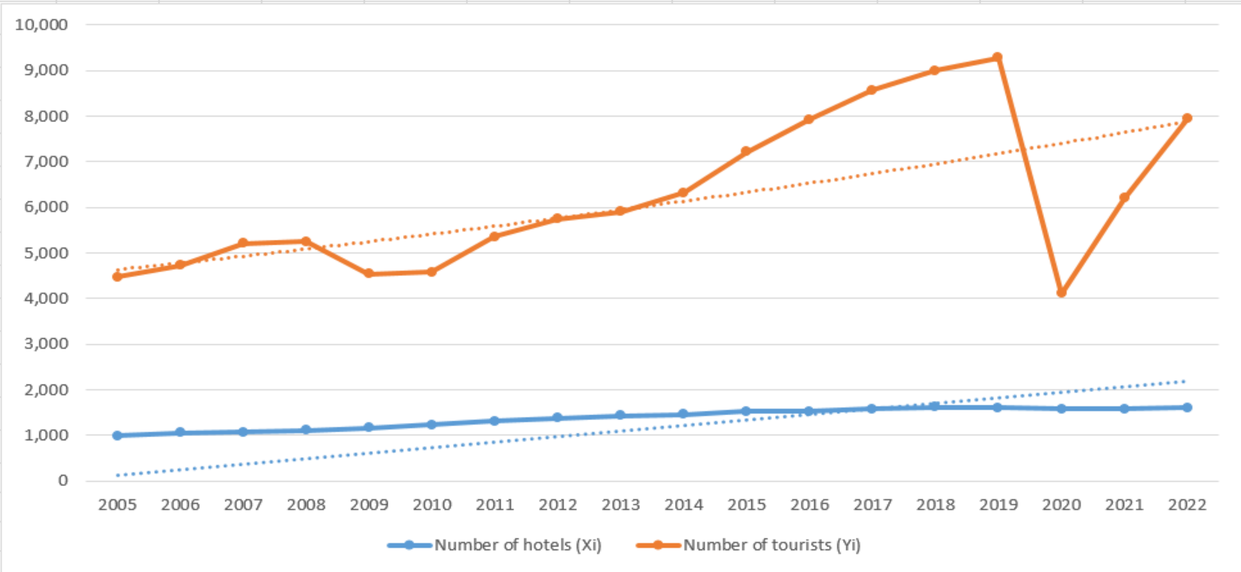

Fig. 3. The graphical representation of the two characteristics
Source: NIS, 2023, http://statistici.insse.ro:8077/tempo-online/#/pages/tables/insse-table [8].

To facilitate the comparison of the two one-dimensional distributions organized into interval groups, we calculated *central tendency indicators* (mean, median, mode), *variability indicators* (variance, squared mean deviation, coefficient of variation), and *skewness indicators* (absolute skewness, relative skewness, interquartile range, skewness coefficient). To assess the strength of the relationship between the two variables (Xi, Yi), we computed the Pearson correlation coefficient, a measure suitable for data with an asymptotic normal or normal distribution.

a. Statistical analysis of frequency distributions - central tendency indicators

Central tendency indicators are a highly significant category of statistical measures employed in the analysis of numerical variables, succinctly capturing what is typical, essential, characteristic, objective, and stable for a set of numerical data. These indicators become more informative about central tendencies as the underlying data becomes more homogeneous [5].

→ ***Mean:*** It is the most commonly used indicator to characterize central tendency and represents the value that, when replacing all terms in a series, does not change their cumulative level. The mean is a typical or central value within a distribution (most frequently used is the arithmetic mean). It is sensitive to extreme values, which can impact the meaning and representativeness of the mean as a central value. For the mean to be representative, the data from which it's calculated should be as homogeneous as possible. When the number of characteristic values coincides with the number of units, the simple arithmetic mean is used. In cases where the same value of the characteristic is recorded for multiple units, the mean is calculated as a weighted arithmetic mean [1]. Since the data from the two analyzed series are not redundant, we have used the simple arithmetic mean:

- for the X_i characteristic:
$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{24{,}815}{18} = 1{,}379$$
- for the Y_i characteristic:
$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{112{,}359{,}392}{18} = 6{,}242{,}188$$

Positional average indicators highlight central tendencies or concentrations of units. Among the positional average indicators, the most commonly used ones are the median and the mode.

→ ***Median:*** It is the central value in an ordered distribution of data, essentially representing the value in an ordered series (ascending/descending) that divides the series into two equal parts, such that 50% of the terms have values lower than the median, and 50% have values higher than the median. The median considers only the position of the terms in the series, remaining unaffected by the magnitude of these values. The median is equal to the central term of the ordered series (ascending/descending) if the series has an odd number of terms. If the series has an even number of terms, the median is equal to the simple arithmetic mean of the two central terms of the ordered series (ascending/ descending) [4].

For the two analyzed series, we have n=18, so the position of the median will be calculated using the formula: (n+1)/2 = 9.5, which means the median lies between the ninth and tenth elements of each series, after they have been arranged in ascending (or descending) order.

- for the X_i characteristic: Me ≈ (1,429 + 1,456)/2 =1,443
- for the Y_i characteristic: Me ≈ (5,908,649 + 6,314,865)/2 = 6,111,757

→ ***Mode (mode of a series):*** It is the value that appears most frequently in a distribution or corresponds to the highest frequency of occurrence.

It can be observed that for both characteristics, the data is not redundant, meaning each value has a single occurrence. However, to determine this indicator, we will use the method of grouping by variation intervals for a frequency distribution series:

Table 2. Data Grouped by Frequency of Occurrence

Xi Characteristic		**Yi Characteristic**	
Intervals	Frequency of occurrence	Intervals	Frequency of occurrence
989 - 1,149	4	4,116,681-5,406,250	8
1,149 - 1,306	2	5,406,249-6,695,818	4
1,306 - 1,463	4	6,695,818-7,985,386	3
1,463 - 1,619	8	7,985,386-9,274,954	3

Source: own contribution.

For characteristic X_i, the median interval size was determined as h_x = 157, and for characteristic Y_i, it was found as h_y = 1,289,568, calculated as the ratio between the absolute amplitude and the number of grouping intervals, which is equal to 4, as shown in Table 2.

The absolute frequency of the median interval for X_i is 4, and the same holds for Y_i. Subsequently, the values for the mode were calculated using the lower limit of the modal interval, the size of the modal interval, the difference between the frequency of the modal interval (Δ_1) and the one before it, and the difference between the frequency of the modal interval and the subsequent one (Δ_2) [3].

- for the X_i characteristic:
$$Mox = x_0 + h_x * \frac{\Delta_1}{\Delta_1 + \Delta_2} =$$
$$= 1{,}306 + 157 * \frac{2}{2+4} \approx 1{,}149$$

- for the Y_i characteristic:

$$Moy = y_0 + h_y * \frac{\Delta_1}{\Delta_1 + \Delta_2} =$$

$$= 5{,}406{,}249 + 1{,}289{,}568 * \frac{4}{4+1} \approx$$

$$\approx 6{,}437{,}905$$

b. Statistical analysis of frequency distributions - variation and skewness indicators

With the help of variation indicators, we can [4], [10]:

-study the representativeness of the mean for a data series;

-assess the degree of homogeneity of the series;

-characterize the degree of variation in a series;

-compare over time and space multiple distribution series for the same characteristic or different characteristics recorded for the same population;

-understand the degree of influence of the grouping factors;

-understand the shape of the frequency distribution through comparison with the normal distribution.

The verification of ***data homogeneity*** involves calculating simple variation indicators (absolute range of variation and relative range of variation), as well as synthetic variation indicators (variance, mean squared deviation, and coefficient of variation).

→ ***The absolute range of variation*** is determined as the difference between the maximum value (x_{max}) and the minimum value (x_{min}) of the characteristic, indicating the widest range of dispersion of characteristic values:

- for the X_i characteristic:

$$A_{abs} = x_{max} - x_{min} = 1{,}616 - 989 = 627$$

- for the Y_i characteristic:

$$A_{abs} = y_{max} - y_{min} = 5{,}158{,}273$$

→ ***The relative range of variation*** is obtained by dividing the absolute range by the mean:

- for the X_i characteristic:

$$A_{rel}{}^{\%} = \frac{x_{max} - x_{min}}{\bar{x}} * 100 = 45.48\%$$

- for the Y_i characteristic:

$$A_{rel}{}^{\%} = \frac{y_{max} - y_{min}}{\bar{y}} * 100 = 82.64\%$$

→ ***The variance*** is calculated as the simple arithmetic mean (for simple series) or weighted mean (for frequency distribution series) of the squared deviations of the series terms from the central tendency (most commonly the arithmetic mean). It is an extremely useful indicator in statistical studies, being used in calculating skewness, kurtosis, and other statistical indicators.

- for the X_i characteristic:

$$\sigma_x{}^2 = \frac{\sum (x_i - \bar{x})^2}{n} = 46{,}053$$

- for the Y_i characteristic:

$$\sigma_y{}^2 = \frac{\sum (y_i - \bar{y})^2}{n} = 2{,}649{,}510{,}644{,}172$$

→ ***Mean squared deviation (standard deviation)*** is calculated as the square root of the arithmetic mean of the squared deviations of the series terms from their mean or as the square root of the variance. The mean squared deviation indicates on average how much the terms of a series deviate from the central tendency (usually the mean).

- for the X_i characteristic:

$$\sigma_x = \sqrt{\sigma_x{}^2} \approx 214.6$$

- for the Y_i characteristic:

$$\sigma_y = \sqrt{\sigma_y{}^2} \approx 1{,}627{,}731.75$$

A notable observation is the high standard deviation for the number of tourists, suggesting a substantial dispersion around the mean.

→ ***The coefficient of variation*** is the most widely used and significant indicator for analyzing variation. It is calculated as the ratio between the standard deviation (either squared or linear) and the mean.

- for the X_i characteristic:

$$v = \frac{\sigma_x}{\bar{x}} * 100 \approx 13.8\%$$

- for the Y_i characteristic:

$$v = \frac{\sigma_y}{\bar{y}} * 100 \approx 26.\%$$

It is noticeable that for both characteristics, the two values indicate a homogeneous dataset. This is evident as both the relative amplitude is below 100%, and the coefficient of variation is under 35%. Nevertheless, the dataset demonstrates greater homogeneity concerning the first characteristic (number of hotels), as indicated by a smaller spread around the mean (45.48%) and a lower intensity of variation (13.8%) compared to the second characteristic (number of tourists).

Determining the **asymmetry** of an empirical distribution series can be done through both graphical methods and by calculating asymmetry indicators. The most commonly used graphical representation for assessing asymmetry is the frequency polygon. However, graphical representations provide only a suggestive insight into the degree of asymmetry, without yielding an exact value to measure it. Evaluating asymmetry involves understanding the standard deviation, the interquartile range, and the coefficient of asymmetry [9].

→ ***The interquartile range (IQR)*** measures the spread of the middle half of the data and is used to assess the dispersion around the median. IQR is calculated as the difference between the third quartile (Q3) and the first quartile (Q1) in a dataset (IQR = Q3 - Q1). Quartiles are values that divide the dataset into four equal parts, with Q1 dividing the lower portion (25% of data) and Q3 dividing the upper portion (75% of data) [11].

To calculate the IQR, the data is first arranged in ascending order, then the positions of the quartiles are identified as follows:

(I)The first quartile (Q1): (n+1)/4 = (18+1)/4 = 4.75, so it falls between the fourth and fifth element,

(II)The second quartile (Q2 or median): (n+1)/2 = (18+1)/2 = 9.5, so it falls between the ninth and tenth element,

(III)The third quartile (Q3): 3*(n+1)/4 = 3*(18 +1)/4 = 14.25, so it falls between the fourteenth and fifteenth element.

- for the X_i characteristic:
  Q1 = 1,196 and Q3 = 1,582,
  so IQR = Q3 - Q1 = 386
- for the Y_i characteristic:
  Q1 = 4,725,448 and Q3 = 8,565,979,
  so IQR = 3,840,531

For the first characteristic, the interquartile range is 386, which means that 50% of the data in the distribution lies within this interval. The relatively large value of the interquartile range for the number of tourists indicates significant variability in the data of this series.

→ ***The skewness coefficient***: measures the asymmetry of data distribution or the deviation from the symmetry of data distribution and can be calculated using Pearson's formula, which involves indicators such as the mean, median, and standard deviation. A positive value of the skewness coefficient indicates positive skewness, to the right, meaning the data is skewed to the left, while a negative value indicates negative skewness, to the left, meaning the data is skewed to the right. A coefficient close to zero indicates an approximately symmetric distribution, meaning perfect symmetry.

- X_i characteristic: $C_{as} = \frac{\bar{x} - Mo}{\sigma_x} \approx 1.15$
- Y_i characteristic: $C_{as} = \frac{\bar{y} - Mo}{\sigma_y} \approx -0.19$

For the first characteristic, the skewness coefficient is approximately 1.15, which suggests a positive asymmetry of the data distribution, meaning there is a longer tail on the left side of the distribution. For the other characteristic, the skewness coefficient has a relatively small, negative value (-0.19), which suggests a slight negative asymmetry. This indicates that the distribution has a longer tail on the left side and is moderately more concentrated on the right side.

All these results provide a more detailed picture of the dispersion and asymmetry in the data concerning the number of hotels and the number of tourists for the analyzed period.

c. Establishing the statistical relationships between variables

To perform a correlation analysis between the two characteristics - the number of hotels (X_i) and the number of tourists (Y_i) - we can calculate the Pearson correlation coefficient as an average of the normalized products of deviations. The correlation coefficient can vary between -1 and 1, where a value of 1 indicates a perfect positive correlation (when an increase in one variable is associated with a proportional increase in the other variable), a value of -1 indicates a perfect negative correlation (when an increase in one variable is associated with a proportional decrease in the other variable), and a value of 0 indicates a lack of linear correlation between variables or that there is no significant linear correlation [6]. The value of the correlation coefficient depends on the shape of the regression line, which is why this indicator is significant for linear correlations and less meaningful for nonlinear correlations (in the latter case, the correlation ratio is used).

For calculating the Pearson correlation coefficient (r), we use the values of the mean ($\bar{x}$, $\bar{y}$) and the standard deviation (σ_X, σ_Y) for each variable [7]:

$$r = \frac{\sum[(x_i - \bar{x}) * (y_i - \bar{y})]}{[(n-1) * \sigma_x * \sigma_y]} \approx 0.854$$

The obtained value for the correlation coefficient indicates a strong positive correlation between the number of hotels and the number of tourists. In other words, there is a strong tendency that when the number of hotels increases, the number of tourists also increases significantly. However, it's important to note that correlation does not necessarily imply a causal relationship. While there is a strong correlation between the number of hotels and the number of tourists, it doesn't necessarily mean that an increase in the number of hotels directly causes an increase in the number of tourists, or vice versa. There are many additional factors and variables that can influence this complex relationship, such as tourism promotion, regional economy, and local events can also influence both variables.

CONCLUSIONS

Current economic theory, which characterizes and analyzes the functioning of economic mechanisms, highlights the multiple interdependencies present in economic activity. Numerous factors, both primary and secondary, essential and non-essential, quantifiable and non-quantifiable, or quantifiable with approximation, influence socioeconomic phenomena and are interconnected in a reciprocal relationship. Tourism represents an important and significant sector within a country's economy, and as such statistical data in tourism provide the opportunity to develop strategies for this sector, and the evaluated indicators serve as sources for monitoring and refining situations and decisions.

Upon examining the two data series in our analysis, it becomes clear that both the global economic crisis spanning from 2008 to 2010 and the SARS-CoV-2 pandemic from 2020 to 2022 exerted significant impacts on the tourism industry. In the short term, events of this nature can lead to substantial disruptions across the broader economy. What holds significance is that the tourism sector holds the potential to rebound and expand following the stabilization of these respective situations. Looking at the longer term, an overarching trend of growth is evident, albeit with yearly fluctuations in the number of tourists arriving at existing lodging establishments. The continuous rise in tourist numbers serves as an impetus for the establishment of new accommodation options, with hotels being the favored choice for lodging structures.

The statistical analysis conducted in this study represents a specific interpretation of the data. A comprehensive and detailed data analysis involves employing multiple statistical methods to arrive at more precise conclusions and predict future trends in the tourism industry. However, as a conclusion from the case study on the number of hotels and tourists accommodated in these lodging structures, a positive correlation can be observed. This suggests a linear relationship, where the growth of accommodation infrastructure (number of hotels) is associated with a significant increase in tourist flow. While causal linkage is not a necessity in statistical correlation, in simpler terms, it can be said that when the number of hotels increases, there is a strong tendency for the number of tourists to also increase.

A crucial aspect for the tourism sector is to embrace the assistance provided by qualified professionals in the field of statistics. These experts can help popularize the need for statistical methodologies and their utility, offering solutions and evidence based on reliable information that supports stakeholders in the sector, including managers who might not naturally perceive these intricacies.

In conclusion, statistics should play a pivotal role in making significant decisions that can influence businesses over both the short and long terms.

REFERENCES

[1]Cristache, S.E., Şerban, D., 2007, Lucrări aplicative de Statistică şi Econometrie (Applied works in statistics and econometrics), ASE Publishing House, Bucharest, pp.191- 216.

[2]Druica, E., Sandu, M., Druica, I., Ianole, R., 2011, Statisitca pe înțelesul tuturor (Statistics made simple), C.H. Beck Publishing House, Bucharest, pp. 50-122.

[3]Gogu, E., 2017, Ghid de formule in statistica (Formula guide in statistics), Universitara Publishing House, Bucharest, pp. 23-28.

[4]Maniu I., 1998, Statistică teoretică. Studii de caz şi aplicaţii (Theoretical statistics. Case studies and applications), Economica Publishing House, Bucharest, pp.98-125.

[5]Maniu, A.I., Voineagu, V., Mitrut, C., 2002, Statistică general (General statistics), Economica Publishing House, Bucharest, pp.189 – 219.

[6]Onicescu, O., Ștefănescu, V., 1979, Elemente de statistică informațională cu aplicații (Elements of informational statistics with applications), Tehnica Publishing House, Bucharest, pp. 78-86.

[7]Statistics Magazine, https://www.revistadestatistica.ro/supliment/wPcontent/uploads/2018, Accessed on 24.07.2023.

[8]Tempo-online databases, http://statistici.insse.ro:8077/tempo-online/tables/insse-table, Accessed on 18.07.2023.

[9]Titan, E., Voineagu, V., Ghita, S., Boboc, C., Tudose, D., 2007, Statistica. Baze teoretice si aplicatii (Statistics. Theoretical foundations and applications), Economica Publishing House, Bucharest, pp.18-56.

[10]Veleanu, I., Hincu, M., 1990, Elemente de statistică generală (Elements of general statistics), Litera Publishing House, Bucharest, pp.43-79.

[11]Wonnacott, T.H., Wonnacott, R.J.,1995, Statistique, Economica, Paris, pp.15-38.