ISSN 2226-3780

UDC 004:330:517
DOI: 10.15587/2312-8372.2018.147613

Lipyanina Kh.

DEVELOPMENT OF INFORMATION TECHNOLOGY OF CORRELATION ANALYSIS OF TOURIST DEMAND

Об'єктом дослідження є процес автоматизації інформаційної технології кореляційного аналізу туристичного попиту на основі когнітивно-статистичного підходу. Одним з найбільш проблемних місць є визначення факторів, що впливають на туристичний попит, тому потрібно розробити інформаційну технологію кореляційного аналізу, яка дозволить визначати фактори, котрі найбільше впливають на туристичний попит.

У роботі обґрунтовано вдосконалення методу кореляційного аналізу туристичного попиту, що представляє розрахунок множинного коефіцієнта кореляції туристичного попиту, відмінною особливістю якого є врахування як якісних, так і кількісних параметрів. Для аналізу туристичного попиту обрано наступні фактори:

– середня заробітна плата на особу в туристичній галузі;
– витрати на туризм;
– кількість колективних засобів розміщування;
– кількість суб'єктів туристичної діяльності;
– кількість рекреацій;
– випуск в основних цінах та випуск за видами економічної діяльності;
– капітальні інвестиції за регіонами;
– транспортне сполучення;
– екологічна ситуація;
– інфраструктура (суб'єктивний показник).

Розроблено блок інформаційної технології моделювання і аналізу для дослідження туристичного попиту, що визначає кореляційну залежність між факторами, що впливають на туристичний попит. Інформаційна технологія розроблена на мові програмування R, пакетом Shiny, яка дає можливість створення інтерактивних веб-додатків, що дає простоту розробленій технології для звичайного користувача.

Виявлено наступні фактори, що впливають на туристичний попит:

– кількість колективних засобів розміщування;
– кількість суб'єктів туристичної діяльності;
– кількість рекреацій;
– рівень інфраструктури на основі суб'єктивної експертизи.

В результаті на основі кореляційного аналізу побудовано модель процесу формування туристичного попиту на основі когнітивно-статистичного методу.

Завдяки цьому забезпечується можливість у подальшому розвивати методологічні засади стратегічного планування розвитку суб'єктів на макро- і мікрорівні, розробляти нормативно-правові, економічні та соціально-політичні механізми гнучкого розвитку туристичних підприємств в окремих регіонах на якісно нових засадах.

Ключові слова: *блок інформаційної технології моделювання, туристичний попит, кореляційний аналіз, мова програмування R.*

1. Introduction

The development of the tourism industry over the past 60 years, as well as the constant increase in revenues from the tourism industry in the world's GDP, are encouraging market participants to pay more and more attention to methods to stimulate demand for tourism services. In particular, the number of international tourist flows increased 43 times: from 25 million people in 1950 to 1087 million people in 2017. Tourism accounts for 6 trillion USD of global investment (4.3 % of the total), which provides 120 million of direct and 176 million of indirect jobs in related industries. The tourism industry is closely related to the market economy and affects the development of 32 sectors [1].

One of the important factors in the development of tourism in Ukraine is tourist demand. According to Ukrinform of Ukraine, last year the trend of increasing the number of entry tourists continued after a two-fold drop in 2014. So, in 2017, Ukraine was visited by 14,600,000 foreigners, which is almost a million more than the previous year. In 2013, Ukraine was visited by 25 million foreign tourists, and in 2014, 12,500,000. Since then, the number of visitors has been growing [2].

Therefore, in this paper it is proposed to use the method of correlation analysis of tourist demand in Ukraine, based on the methods of subjective examination of quality indicators and their statistical analysis. These facts determine the relevance of the study, and an information technology has been developed for the correlation analysis of tourist

demand that will allow the specialists and interested users to quickly solve various individual tasks.

2. The object of research and its technological audit

The object of research is the process of automating information technology of the correlation analysis of tourist demand on the basis of a cognitive-statistical approach. Demand is one of the most important factors stimulating the development of tourism. Demand is closely related to the supply of any tourist product – a tourist service, a product, a tourist and recreational resource, infrastructure and tourist conditions.

One of the most problematic places is the identification of factors affecting tourist demand. For what and need to develop information technology correlation analysis, which will determine the factors that most affect tourism demand.

At a certain stage of development, almost every subject of the tourism industry is faced with the need to reorganize the tourism business in response to changes in the external environment. The current status of information-computer technology (ICT) and automation is one of the most significant factors contributing to the development in the tourism business. The pace of activity is accelerating and in order to keep revenues at an appropriate level with the loyalty of customers (tourists) and travel agency employees is not the only factor in increasing the competitiveness of an organization.

To do this, it is necessary to develop a program that automates the process of determining the tourist demand takes place in Ukraine, which will give an opportunity to assess the tourist demand to any user.

When creating this software, orientation was carried out on a modern Web interface. This information technology is implemented using the *R* programming language with the Shiny package, which allows using various data storages to import information. Shiny is a package that allows to create interactive web applications directly in *R*.

To develop information technology for modeling and analyzing tourist demand on the basis of a cognitive-statistical approach, it is decided to automate the following elements: information collection, dependency building, forecasting (Fig. 1).

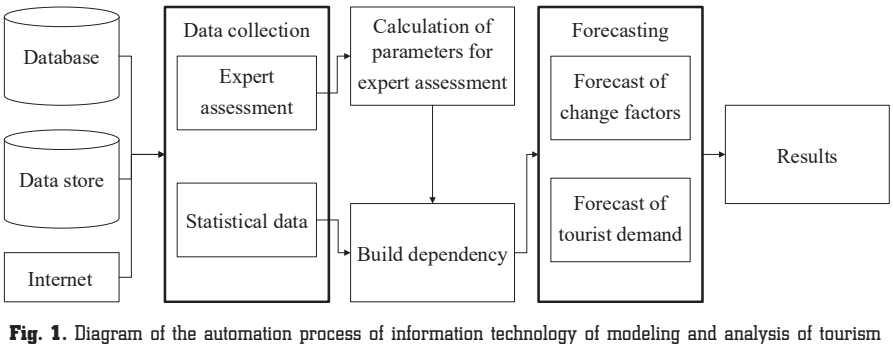

Fig. 1. Diagram of the automation process of information technology of modeling and analysis of tourism demand on the basis of a cognitive-statistical approach

To build an information technology for modeling and analyzing tourism demand on the basis of a cognitive-statistical approach, the following method is proposed:

Step 1. Import data from a database in .csv, .xls, .dat, .sav, .txt format, or directly enter data into the software environment.

Step 2. Preliminary analysis of data, recovery of missing values, preparation of the time series vector for further analysis.

Step 3. Calculation of the parameters of expert assessment by the factor of the level of infrastructure in the region.

Step 4. Calculation of the correlation dependence between the factors influencing tourist demand.

Step 5. Building a regression model of tourist demand in Ukraine based on a multifactor model.

Step 6. Construction of the predicted value of indicators of tourist demand by means of ARIMA-modeling.

3. The aim and objectives of research

The aim of research is development of an information technology for the correlation analysis of tourist demand, a distinctive feature of which is to take into account both qualitative and quantitative parameters. This will highlight the main factors that influence tourist demand.

To achieve this aim, it is necessary to perform the following objectives:

1. Improve the method of correlation analysis of tourist demand.

2. Develop a block of information technology, determines the correlation dependencies between the factors influencing tourist demand.

3. Build on the basis of the correlation analysis model of the process of formation of tourist demand.

4. Research of existing solutions of the problem

Decision support systems are used to manage the tourist area, in which information technologies for modeling and analyzing tourist demand play a significant role. At present, for solving this problem, two main approaches are used: statistical and cognitive. In the statistical approach, using the method of regression estimation of dependencies, however, before this, determine the correlation relationship between factors that may affect the object under study.

Significant contribution to the development of theoretical and applied issues of modeling socio-economic processes that affect the dynamics of tourist flows are made by many scientists. In particular, the authors of the study [3] built an econometric model of tourist demand in the developed tourist market in order to improve the tourist product of Greece. In [4], study of assessing the economic determinants of international demand for tourism services in Spain is conducted based on econometric models, taking into account such factors as: real income per capita, exchange rates and real prices for Spanish tourist services.

Applied research [5] of tourist demand in France with the help of an econometric model shows a positive relationship between the costs of tourism and the production of a country's GDP, as well as the negative correlation between tourism costs and prices. The authors of [6] develop an econometric model of the demand for tourism

ISSN 2226-3780

according to the tourist flows in Western Europe and proposed prediction models. Spatial econometric approach to modeling tourist flows is presented in the study [7]. The study of the tourism industry with the use of economic and mathematical modeling was carried out in [8, 9], which identified the main factors affecting the activities of enterprises in the tourism sector. Using the correlation-regression analysis, the influence of the proposed factors on the functioning and development of tourism enterprises is evaluated, the most significant factors are selected and a number of econometric models are proposed that reflect these dependencies. In [10], the territorial organization of the markets for inbound tourism in Ukraine is studied, both using factor analysis and based on the developed cartographic models. The dynamics of tourist flows in Ukraine and the identification of factors influencing their change using econometric methods and models are studied by the authors of [11]. Using the correlation-regression analysis in [12], the author assess the impact of economic indicators on the value of the outbound tourist flow of the Transcarpathian region (Ukraine).

Existing information technologies for modeling and analyzing tourist demand (COMFAR, PROPSPIN, Alt-Invest, Project Expert, Microsoft Project, Time Line, Primavera, Vortex, Forecast Expert) do not provide the necessary quality practices, which significantly reduces the effectiveness of decision support systems.

Based on the analysis, it can be concluded that many scientists have developed a wide range of theoretical and practical aspects of modeling the socio-economic situation in various areas of the country. However, the features of information technology in relation to the correlation analysis of tourist demand, taking into account the factors pushing, attracting and repelling, are ignored.

5. Methods of research

In this paper, let's construct a block of information technology (IT) modeling and analysis of tourist demand on the basis of a cognitive-statistical approach, which determines the correlation dependence between the factors influencing tourist demand.

The demand for tourism is expressed by the number of arriving tourists or the expenses incurred in the country. In the tourism demand model, macroeconomic indicators are taken into account, such as income in countries, the price of tourism in Ukraine, the cost of transport, and the number of collective accommodation facilities.

Tourist demand in Ukraine is characterized on the basis of the following factors, based on scientific research [3–5]:

$$Y = F(S, V_{TR}, P, R, K, C, I, T, Z, N), \quad (1)$$

where Y – flow of tourists; S – average salary per person in the tourism industry; V_{TR} – tourism expenses; R – the number of collective accommodation facilities; P – the number of subjects of tourist activity; K – the number of recreation; C – release in basic prices and release by types of economic activity; I – capital investment by region; T – transport service; Z – environmental situation; N – infrastructure (subjective indicator).

To calculate the multiple correlation coefficient of tourist demand, a distinctive feature of which is to take into account both qualitative and quantitative parameters, it is necessary:

Step 1: determine the factors of the model (1), most significantly affect the tourist demand Y.

Step 2: calculate the sum of the values of the factors of the model (1). Calculate the arithmetic average of the factors, calculate the standard deviations using the standard deviation formula for each factor.

Step 3: build a matrix (A) of paired correlation coefficients r_{ij}, $i = \overline{1,k}$ between the signs of model factors (1) using the Pearson correlation coefficient r_{xy}:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

where x_i i y_i – the values of the variable factors of the model (1); $\bar{x}$ i $\bar{y}$ – the average value of the variables of the model (1); n – sample size:

$$A = \begin{bmatrix} 1 & r_{SV_{TR}} & r_{SP} & r_{SR} & r_{SK} & r_{SC} & r_{SI} & r_{ST} & r_{SN} \\ r_{V_{TR}S} & 1 & r_{V_{TR}P} & r_{V_{TR}R} & r_{V_{TR}K} & r_{V_{TR}C} & r_{V_{TR}I} & r_{V_{TR}T} & r_{V_{TR}N} \\ r_{PS} & r_{P\,V_{TR}} & 1 & r_{PR} & r_{PK} & r_{PC} & r_{PI} & r_{PT} & r_{PN} \\ r_{RS} & r_{R\,V_{TR}} & r_{RP} & 1 & r_{RK} & r_{RC} & r_{RI} & r_{RT} & r_{RN} \\ r_{KS} & r_{K\,V_{TR}} & r_{KP} & r_{KR} & 1 & r_{KC} & r_{KI} & r_{KT} & r_{KN} \\ r_{CS} & r_{C\,V_{TR}} & r_{CP} & r_{CR} & r_{CK} & 1 & r_{CI} & r_{CT} & r_{CN} \\ r_{IS} & r_{I\,V_{TR}} & r_{IP} & r_{IR} & r_{IK} & r_{IC} & 1 & r_{IT} & r_{IN} \\ r_{TS} & r_{TV_{TR}} & r_{TP} & r_{TR} & r_{TK} & r_{TC} & r_{TI} & 1 & r_{IN} \\ r_{NS} & r_{NV_{TR}} & r_{NP} & r_{NR} & r_{NK} & r_{NC} & r_{NI} & r_{NT} & 1 \end{bmatrix}.$$

Step 4: estimate the density of the relationship between the dependent variable Y and the independent variables of the model factors (1) in accordance with the table of values of the correlation coefficient values.

There are several methods for determining the level of correlation coefficient. The most famous is the least squares method. However, this rather time-consuming calculation can be successfully replaced using functional-statistical dependencies in the R programming language.

6. Research results

Before beginning the correlation analysis, it is necessary to visualize the data. Statistical indicators are taken from the State Statistics Service of Ukraine [2]. The level of infrastructure is determined on the basis of a subjective examination [13].

The dependencies between the indicators are quite close, because they are modeled as linear. In a situation where curved patterns appear in the scattering regions, let's deal with a nonlinear association between two variables. Therefore, the graph in Fig. 2 shows that all dependencies have a linear relationship.

While it seems that the distribution is normal. And now let's check formally. One of the statistical criteria for checking the normality of the distribution of data is the Shapiro-Wilk test. Using this criterion, the null hypothesis is verified, which is that the data is distributed normally.

If, by the Shapiro-Wilk criterion, $P\text{-}value > 0.05$, then the probability that the hypothesis is correct is zero. If the significance level is less than 5 % (0.05), there is reason to reject the null hypothesis that the data are normally distributed. The turnout has a normal distribution.

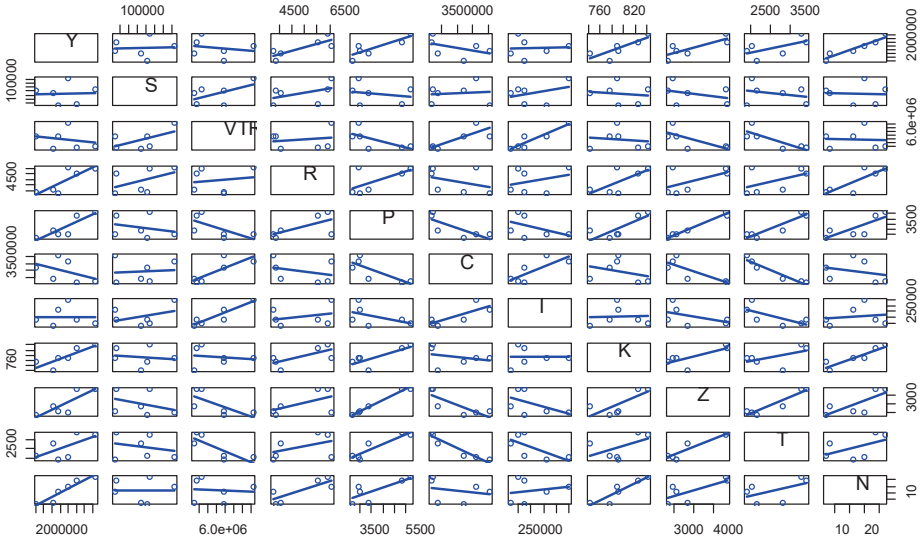

Fig. 2. Visualization of data affecting tourism demand

As can be seen from the results in the Table 1, *P-value* values are larger than a predetermined level of 0.05 values. This means that the distribution of data is not significantly different from the normal distribution. In other words, it is possible to assume that the distribution is normal.

Table 1

Significance by Shapiro-Wilk criterion

Indicator	Shapiro-Wilk criterion	*P-value*
Y – flow of tourists	0.99	0.09
S – average salary per person in the tourism industry	0.91	0.06
V_{TR} – tourism expenses	0.88	0.28
R – the number of collective accommodation facilities	0.82	0.09
P – the number of subjects of tourist activity	0.83	0.11
K – the number of recreation	0.88	0.25
C – release in basic prices and release by types of economic activity	0.88	0.28
I – capital investment by region	0.96	0.04
T – transport service	0.82	0.08
Z – environmental situation	0.86	0.21
N – infrastructure (subjective indicator)	0.92	0.02

The correlation matrix is used to study the relationship between several variables. The result is a table containing the correlation coefficients between each variable and the others. The correlation matrix is always symmetric (the correlation coefficient between X and Y variables is equal to the correlation coefficient between Y and X variables), and the main diagonal of such a matrix is 1 (the correlation of the variable with itself is 1).

A correlation analysis of the process of formation of tourist demand has been carried out (Fig. 3). There are several methods for determining the level of correlation coefficient. The most famous is the least squares method. However, this rather time-consuming calculation can be successfully replaced by the use of functional-statistical dependencies in the *R* programming language, which makes it possible to reduce the search for correlation coefficients to several minutes.

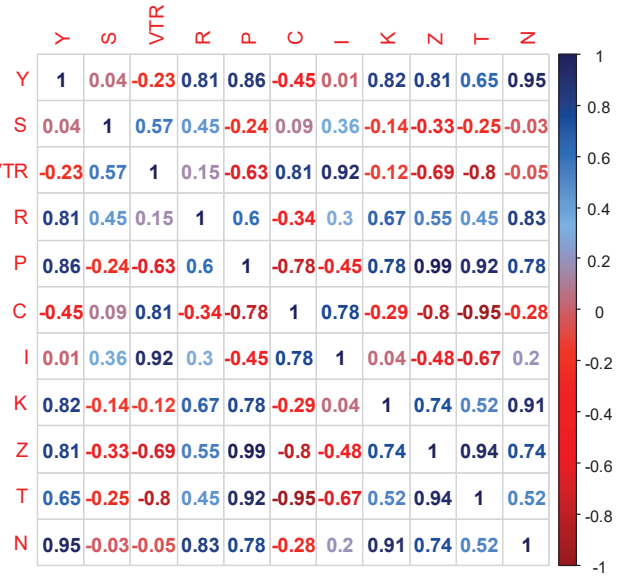

Fig. 3. Matrix of tourist demand correlations

The calculation of the correlation matrix allows to conclude that there is a significant relationship between the effective index (Y) and the factor values, in addition to several factors (S, VTR, C, I and T). More detailed information is presented in Table 2.

Based on the correlation analysis, a model of the process of formation of tourist demand has been built, which will look like this:

$$Y = F(P,R,K,N),$$

where Y – the flow of tourists; R – the number of collective accommodation facilities; P – the number of subjects of tourist activity; K – the number of recreation; N – the level of infrastructure based on subjective expertise.

Fig. 4 shows the program code in the *R* block of information technology modeling and analysis for the study of tourist demand determines the correlation dependence between the factors influencing tourist demand.

ISSN 2226-3780

Table 2

Estimation of the relationship density between the dependent variable by *Y* independent variables

Effective index (dependent variable *y*)	Factor score (independent variable *x*)	The value of the pair correlation coefficient (*r*)	Characteristic (density) of linear relationship	Relationship character
Y – the flow of tourists	*R* – the number of collective accommodation facilities	0.8137670	strong	direct
	S – the average salary per person in the tourism industry	0.04219225	practically absent	direct
	V_{TR} – tourism expenses	–0.2271771	practically absent	indirect
	P – the number of subjects of tourist activity	0.8613155	strong	direct
	Z – ecology	0.8080230	strong	direct
	N – level of infrastructure based on subjective expertise	0.94793356	very strong	direct
	K – the number of recreation	0.8180624	strong	direct
	C – release in basic prices and release by type of economic activity	–0.44915574	weak	indirect
	I – capital investment by region	0.006036328	practically absent	direct
	T – transport connection	0.6452539	perceptible	direct

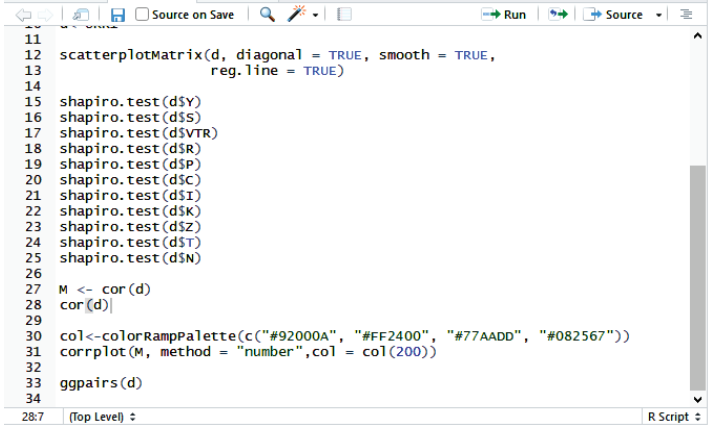

Fig. 4. Program code in *R*

A visual representation of the information technology modeling and analysis unit for the study of tourist demand determines the correlation dependencies between the factors influencing tourist demand, are presented in Fig. 5.

The proposed information system provides a definition of tourist demand, aimed at improving the work of the tourism industry and are implemented in the Department of Tourism and Resorts of the Ternopil Regional State Administration (Ukraine) and the Travel Agency «Everywhere Together» (Ternopil, Ukraine).

7. SWOT analysis of research results

Strengths. The proposed description allows to develop a block of information technology modeling and analysis for the study of tourist demand determines the correlation between the factors influencing tourist demand. Due to this, it will increase the efficiency of the analysis by reducing the level of uncertainty about the compliance of competence and personal qualities of the performers to the degree of complexity of the project.

Weaknesses. The proposed approach, since it is based on expert assessment, carries with it the organic shortcomings of these methods, namely, some subjectivity, although its influence on the formation of decisions by the expert commission is significantly reduced.

Opportunities. In the future, it seems appropriate to carry out modeling based on the results obtained and to develop a full-fledged technology for modeling and analyzing tourist demand on the basis of a cognitive-statistical approach. The proposed information system provides a definition of factors affecting tourism demand, which makes it possible to predict the number of tourists, and accordingly, predict the future profit of a tourist enterprise.

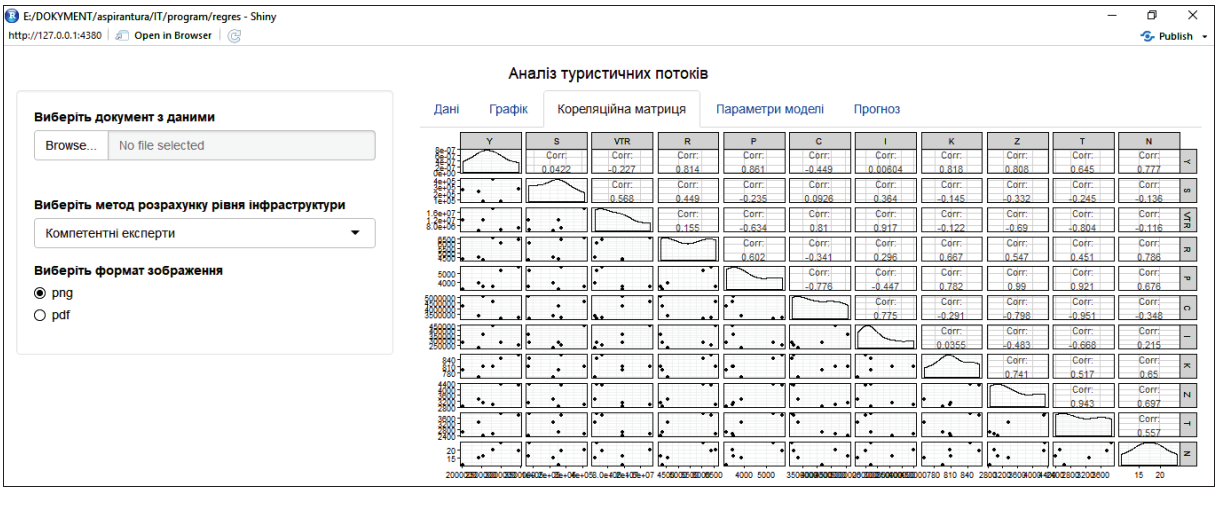

Fig. 5. Window of the correlation matrix representation in the form of a graph

Threats. The proposed description when stored in the information technology can potentially be destroyed in whole or in part. The reason for this is biased or impartial actions of users and personnel serving the information technology of the travel agency, and therefore requires the implementation of special security measures.

8. Conclusions

1. The method of correlation analysis of tourist demand is improved, which is the calculation of the multiple correlation coefficient of tourist demand, a distinctive feature of which is to take into account both qualitative and quantitative parameters. The following factors were chosen for the analysis of tourist demand:

– the average wage per person in the tourism industry;
– tourism expenses;
– the number of collective accommodation facilities;
– the number of subjects of tourist activity;
– the number of recreation;
– release in basic prices and release by types of economic activity;
– capital investment by region;
– transport connection;
– ecological situation;
– infrastructure (subjective indicator).

2. A block of information technology for modeling and analysis is developed for the study of tourist demand which determines the correlation dependence between the factors influencing tourist demand. Information technology is developed in the *R* programming language, by the Shiny package, which enables the creation of interactive web applications and the simplicity of the developed technology for the average user.

3. The following factors affecting tourist demand are identified:

– the number of collective accommodation facilities;
– the number of subjects of tourist activity;
– the number of recreation;
– level of infrastructure based on subjective expertise.

As a result, based on the correlation analysis, a model of the process of formation of tourist demand on the basis of the cognitive-statistical method is built.

References

1. UNWTO Turism Highlights. URL: https://www.e-unwto.org/doi/pdf/10.18111/9789284419876
2. Derzhavna sluzhba statystyky Ukrainy. URL: http://www.ukrstat.gov.ua/
3. Dritsakis N., Athanasiadis S. An Econometric Model of Tourist Demand // Journal of Hospitality & Leisure Marketing. 2008. Vol. 7, Issue 2. P. 39–49. doi: http://doi.org/10.1300/j150v07n02_03
4. Garin-Munoz T., Amaral T. P. An econometric model for international tourism flows to Spain // Applied Economics Letters. 2000. Vol. 7, Issue 8. P. 525–529. doi: http://doi.org/10.1080/13504850050033319
5. An Econometric Model of Tourisn Demand in France / Botti L. et. al. // Tourismos: an International Multidisciplinary Journal of Tourism. 2007. Vol. 2, Issue 1. P. 115–126.
6. Witt S. F., Martin C. A. Econometric Models for Forecasting International Tourism Demand // Journal of Travel Research. 1987. Vol. 25, Issue 3. P. 23–30. doi: http://doi.org/10.1177/004728758702500306
7. Yang Y., Wong K. K. F. A Spatial Econometric Approach to Model Spillover Effects in Tourism Flows // Journal of Travel Research. 2012. Vol. 51, Issue 6. P. 768–778. doi: http://doi.org/10.1177/0047287512437855
8. Kalchenko O. M. Otsinka vplyvu faktoriv rozvytku pidpryiemstv turystychnoi sfery // Naukovyi visnyk Chernihivskoho derzhavnoho instytutu ekonomiky i upravlinnia. Seriia: Ekonomika. 2013. Vol. 3. P. 94–101.
9. Balashova R., Ivchenko L. Metodychni zasady analizu ta prohnozuvannia rynku turystychnykh posluh v Ukraini z vykorystanniam matematychnoho modeliuvannia // Ekonomika. 2011. Vol. 3 (110). P. 3–9.
10. Beidyk O. O., Novosad N. O. Faktornyi analiz formuvannia potokiv viznoho turyzmu Ukrainy // Ukrainskyi heohrafichnyi zhurnal. 2012. Vol. 1. P. 44–49.
11. Barna M. Y., Myronov Y. B. Econometric modelling of tourist flows dynamics // Scientific bulletin of Polissia. 2017. Vol. 1, Issue 4 (12). P. 165–170. doi: http://doi.org/10.25140/2410-9576-2017-1-4(12)-165-170
12. Morokhovych V. S. Ekonomiko-matematychne modeliuvannia turystychnykh potokiv zakarpatskoi oblasti // Naukovyi visnyk Uzhhorodskoho universytetu. 2017. Vol. 1. P. 143–146.
13. Kuvaieva V. I., Lipyanina Kh. V., Boltonkov V. O. Processing expert information in the context of collective assessment of a tourist infrastructure // Innovative Technologies and Scientific Solutions for Industries. 2018. Vol. 5, Issue 3 (5). P. 35–43. doi: http://doi.org/10.30837/2522-9818.2018.5.035

Lipyanina Khrystyna, *Lecturer, Department of Economic Cybernetics and Informatics, Ternopil National Economic University, Ukraine, e-mail: xrustya.com@gmail.com, ORCID: http://orcid.org/0000-0002-2441-6292*