International Journal of Manufacturing Economics and Management
ISSN 2784-1278 ISSN-L 2784-1278, Vol. III, No. 1 / 2023
https://doi.org/10.54684/ijmem.2023.3.1.24

STUDY OF MATHEMATICAL MODELING OF THE FORMATION OF TRANSPORT LOGISTICS TERMINALS

Ziyoda Mukhamedova, Gulshan Ibragimova, Sherzod Fayzibaev

"Tashkent State Transport University, Department of Transport and cargo systems
"Temiryulchilar" str., Tashkent, Uzbekistan, 100167

Corresponding author: Ziyoda Mukhamedova, mziyoda@mail.ru

Abstract: The article is devoted to cluster analysis of the regions of the potential location of logistics terminals. The subjects of Uzbekistan are divided into four clusters depending on the level of their socio-economic, infrastructural development, geographical location and the volume of transport work performed. A statistical model is developed that includes a set of equations and dependency functions of such socio-economic factors in the formation of the logistics infrastructure of the regions as average per capita income of the population, the gross regional product (GRP), specific volume of trade and industrial production, exports and imports, the level of investment, the number of the economically active population, the volume of fixed assets and transport services, as well as indices of transport work - the physical volume of transportation by road and rail. The dependence of the volume of transport services in the regions of the Republic of Uzbekistan on the income of the population, the gross regional product, the density of roads, the volume of exports and belonging to clusters was established. As a result, the regions of Uzbekistan, in terms of the location of freight infrastructure on the railways in order to develop potential transport and logistics clusters, were grouped depending on the indices of socio-economic and infrastructure development.
Key words: railway transport, statistical model, logistics terminal, distribution center, shortest distance, cluster, cluster analysis, correlation analysis, regression.

1. INTRODUCTION

When solving the problem of the optimal location of transport logistics terminals on the railway network in order to provide logistics capacities to existing and emerging industrial processing clusters, it is required to find such a location of distribution centers or sites relative to their suppliers and consumers, at which a certain target function of total logistics costs reaches its minimum values with a comprehensive account of all significant influencing factors. Mathematically, this problem serves the purpose of multi-objective optimization in the presence of a system of restrictions. The solutions of such problems were studied in detail in the works of Mukhamedova Z.G. (Mukhamedova, 2016; Mukhamedova, 2020;), Ibragimova G.R. (Mukhamedova and Ibragimova, 2022), Fayzibaev (Fayzibaev et al., 2022), Suyunbaev Sh.M. (Suyunbaev et al., 2023), Boltaev S.T. (Saburov et al., 2022) and etc.

2. RESEARCH METHODS

2.1 Task stages

The problem of the optimal location of the logistics terminals with the possibility of loading and warehousing should take into account such basic factors as the distance between the warehouse and suppliers and consumers, the volume of cargo transported, transport rates and the time of delivery of goods from suppliers to the warehouse and from the warehouse to consumers, and be solved by determining the coordinates (x, y) of the logistics terminal so that the logistics costs equal to the sum of the products of the distances from suppliers to the logistics terminal and from the logistics terminal to the destination, which has coordinates (x_i, y_i), and the volumes of cargo transported (need or demand) were minimal, as shown by formula 1:

$$P = \sum_{i=1}^{n} Q_i d_i \to \min \tag{1}$$

where: d_i is the distance from the logistics terminal to the i-th supplier or to the destination point (i=1, 2, … n).
If the task is enlarged to the level of location of distribution logistics centers in the country, then the model development procedure should be divided into several stages (Sai, 2015).

2.2 First stage
To determine the level of dependence between the factors influencing the location of logistics centers; the values of dependent (endogenous) variables y_1, y_2, … y_m are estimated depending on the values of independent (exogenous) variables x_1, x_2, … x_k taking into account the influence of random components ε_1, ε_2, … ε_m that cannot be measured (the influence of factors and random errors in the measurement of indicators is not included in the model).
The exogenous or independent variables of the statistical model include climatic zones, availability of roads and railways, availability of transport corridors in the region, and population.
The remaining variables are defined as dependent (endogenous) ones.
Thus, a data array that statistically describes and allows assessing the change in time and space in the system of factors for the location of logistics centers consists of P factors for N regions of Uzbekistan over time intervals T (Ivanova et al., 2012).
A preliminary analysis showed that the logistics centers available in Uzbekistan are located in regions with high trade and industry development, dense population and a favorable geographical location in relation to land transport corridors (Makhkamov et al., 2022).
The assessment of the system of factors for the location of logistics centers was conducted based on a space-time sample (17 indices for 13 regions (the city of Tashkent was excluded from the sample due to sufficient availability of logistics centers), the study period was 2010-2022).
The indices are included in the statistical model with different regression coefficients, the values of which depend on whether the region belongs to a particular cluster. To obtain more accurate regression coefficients and take into account the influence of the region of Uzbekistan belonging to a certain cluster, dummy variables were introduced:

$$r_1 = \begin{cases} 1 - \text{the subject belongs to the first cluster;} \\ 0 - \text{others;} \end{cases}$$
$$r_2 = \begin{cases} 1 - \text{the subject belongs to the first cluster;} \\ 0 - \text{others;} \end{cases}$$
$$r_3 = \begin{cases} 1 - \text{the subject belongs to the first cluster;} \\ 0 - \text{others;} \end{cases}$$
$$r_4 = \begin{cases} 1 - \text{the subject belongs to the first cluster;} \\ 0 - \text{others;} \end{cases} \qquad (2)$$

The model of dependence of the *i*-th factor has the form:

$$Y_i(X) = f(Y, X_1, X_2, \dots, X_p, r_1, r_2, r_3, r_4) + \varepsilon_i \qquad (3)$$

where: ε_i is the random component.
The form of dependence of the *i*-th index (Y_i) on X has the form:

$$Y_i(X) = \beta_0 + \sum \alpha_k * r_k + \sum \beta_j * X_j + \varepsilon_j \qquad (4)$$

where: X_j is the value of j variable that has the strongest influence on i index;
β_0 and β_j are the unknown coefficients of the regression equation;

r_k is a dummy variable, where k is the cluster number;

r_k is the regression coefficient for a dummy variable.

To estimate the parameters of the regression equation, multiple regression analysis is used, since the number of variables on which the j-th index depends is more than one (Sai et al., 2010).

2.3 Second stage

Cluster analysis is performed using various software systems (for example, Statistica) to group multidimensional objects and present the results of individual observations by points of appropriate geometric space, followed by the selection of groups as clusters. Objects included in a certain cluster have similar properties.

Similarity with other objects is defined as the corresponding distance between objects in space, that is, the value of d_{ab} satisfies the axioms (Bureeva, 2007):

$$A1.\ \mathrm{d}_{ab} > 0,\ \mathrm{d}_{ab} = 0,$$
$$A2.\ \mathrm{d}_{ab} = \mathrm{d}_{ba},$$
$$A3.\ \mathrm{d}_{ab} + \mathrm{d}_{bc} \geq d_{ac}. \quad (5)$$

It is appropriate to use the geometric distance between two points in a multidimensional space (the Euclidean distance) as a measure of distance:

$$d_{ab} = \sqrt{\sum_{i=1}^{K} (X_{ia} - X_{ib})^2} \quad (6)$$

where: X_{ia}, X_{ib} are the values of i - th attribute of the a-th (b-th) object (i = 1, 2, ..., k, a, b = 1, 2, … n).

Cluster analysis was conducted using the Excel program. All indices are normalized by the ratio of the difference between the initial and arithmetic mean values of the indices to the standard deviation (in the developed model, this tab is called "Cluster analysis").

Next, the resulting data is selected and, using the Insert-PivotTable algorithm, is opened in a new sheet (in the developed model, this tab is called "Shortest Distances"). Here, the "regions" parameter is transferred to the columns, and the "region number" parameter (No) is transferred to the rows according to the selected value "Volume of freight rail transportation" ($V_{пер.}$ railway, million tons) with the field parameters set to the value "Maximum by field" (Figure 1).

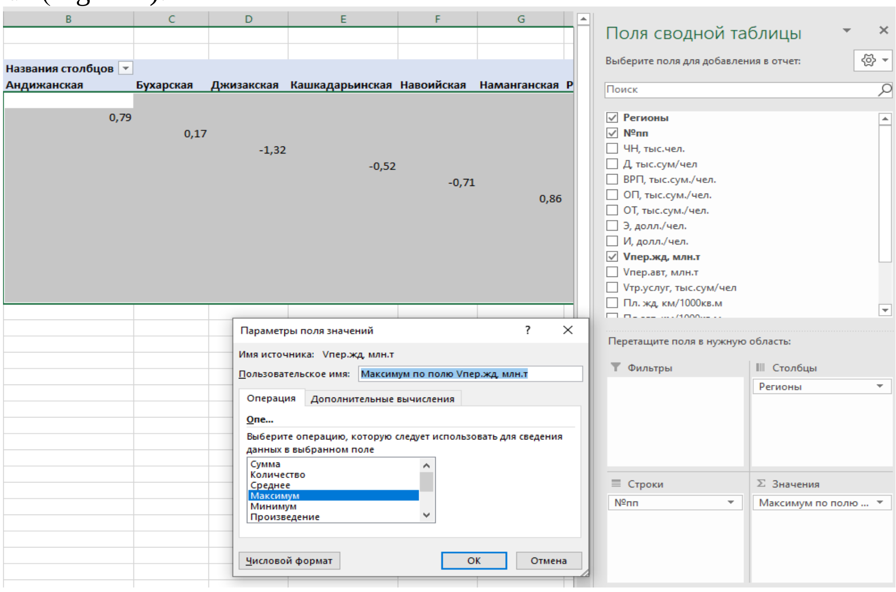

Fig. 1. Setting parameters to determine the shortest distance to the center of the cluster in the cluster analysis of the regions of Uzbekistan

Having previously chosen the value of 4 as k (the number of clusters), formulas are set for the resulting data arrays, highlighting such factors as the distance to each cluster, the shortest distance, the cluster number, and the number of regions in each cluster. The target function is set (the sum of the shortest distances). The “Search for a solution” tool is used in the “Data” tab, where it is selected as the value restriction with the search for the smallest value. The “Evolutionary search for a solution” method is used. The resulting data is split using the “Color Scale” tool in the “Conditional Formatting” tab; the regions of each cluster are classified by data commonality, as shown in Figure 2.

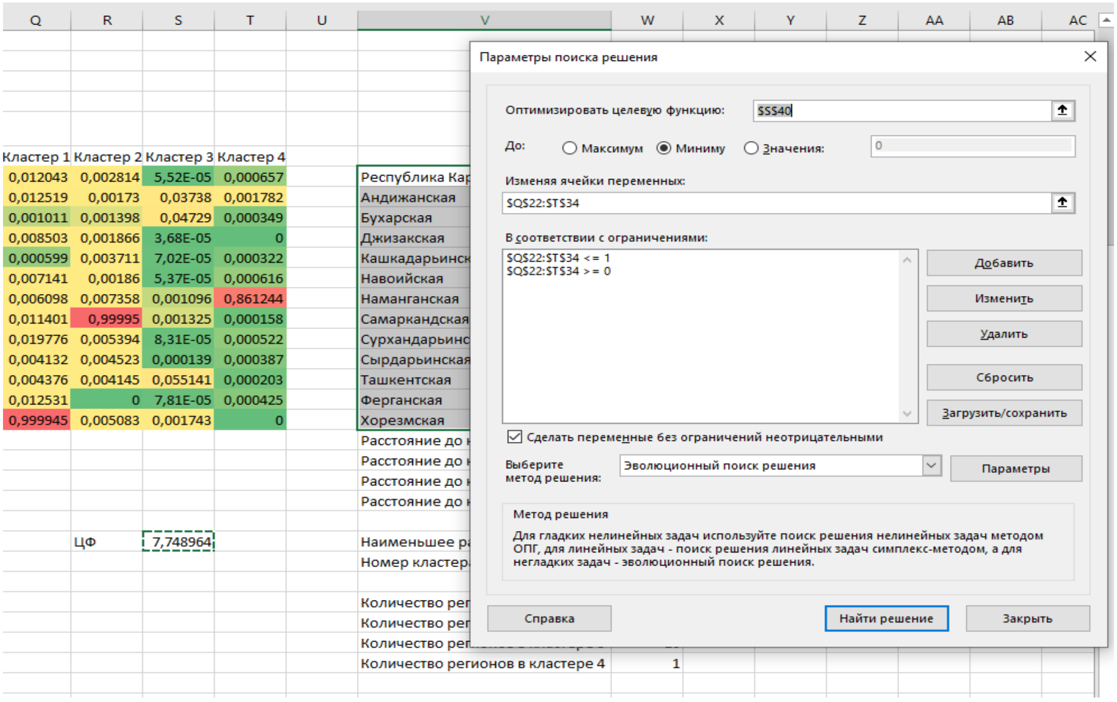

Fig.2. Setting the parameters of the “Search for a Solution” tool in the cluster analysis of the regions of Uzbekistan

As a result of the cluster analysis, 4 clusters were identified:
1st cluster - Khorezm region;
2nd cluster - Samarkand region;
3rd cluster - Republic of Karakalpakstan, Andijan region, Bukhara region, Jizzakh region, Kashkadarya region, Navoi region, Surkhandarya region, Syrdarya region, Tashkent region, Fergana region.
4th cluster - Namangan region.

2.4 Third stage

Paired correlation analysis was used to determine the presence and type of relationship between the studied factors. For correlation coefficient $r < 0.25$ - the correlation is weak, for $0.25 < r < 0.7$, it is moderate, for $r > 0.75$, it is strong (the relationship between the variables is close to linear).
The Pearson correlation coefficient is determined by the following formula:

$$r_{xy} = \frac{\sum_{j=1}^{n} \left[\left(x_j - \bar{x} \right) \cdot \left(y_j - \bar{y} \right) \right]}{(n-1) \cdot \sigma_x \cdot \sigma_y} \qquad (7)$$

where x_j and y_j are compared factors j of the subject;
n is the number of compared observations;
σ_x and σ_y are standard deviations in comparable series.
The analysis was conducted using the Data Analysis tool in Excel using the obtained correlation matrix.

3. RESEARCH RESULTS AND DISCUSSION

According to the results of the calculations, the volumes of transport services are correlated with population income, GRP, industrial production, trade, exports and imports.

Correlations in this case are not considered causal, they show that with a change in one index, other indices also change; this can be influenced by a large number of different reasons. There is an obvious need for additional statistical studies of the type and levels of dependencies between indices. Dependency graphs are shown in Figures 3-15.

In Figure 3, the dependence of the volume of transport services per capita on the population is moderately inverse (correlation coefficient is minus 0.31679). The higher the population in the region, the lower the indicator.

As seen in Figure 4, between the specific volume of transport services and the average per capita income of the population there is a moderate correlation (correlation coefficient is 0.67106). The higher the income of the population, the greater the volume of transport services.

At the same time, in Figure 5, the dependence of the volume of transport services per capita on the volume of GRP is moderately positive (correlation coefficient is 0.691116). The lower the GRP of the region, the higher the volume of specific transport services.

In Figure 6, the dependence of the volume of transport services per capita on the volume of industrial production is moderate (correlation coefficient is 0.65673). The volume of transport services is growing with the growth in industrial production.

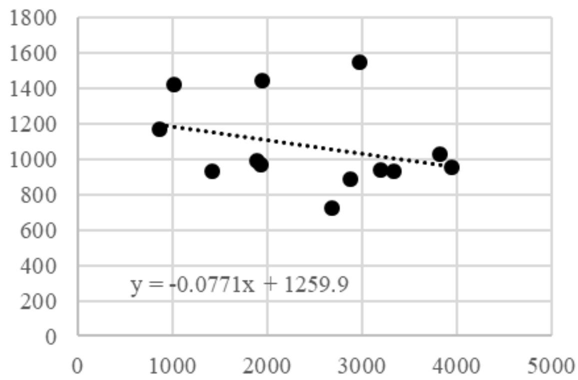

Fig. 3. Dependence of the volume of transport services per capita on the population

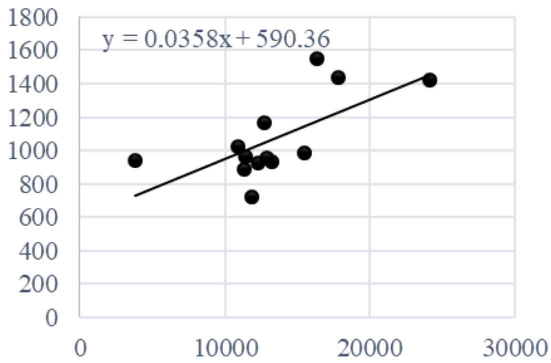

Fig. 4. Dependence of the volume of transport services per capita on the average income of the population

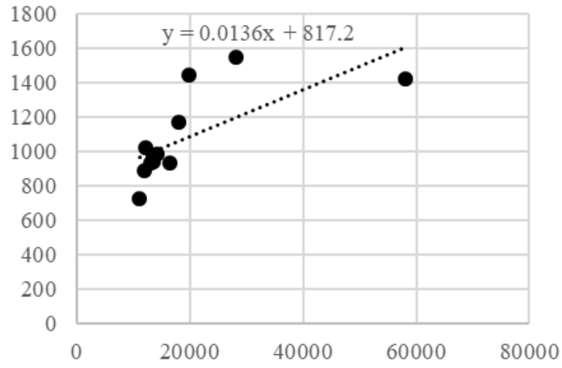

Fig. 5. Dependence of the volume of transport services per capita on the gross regional product

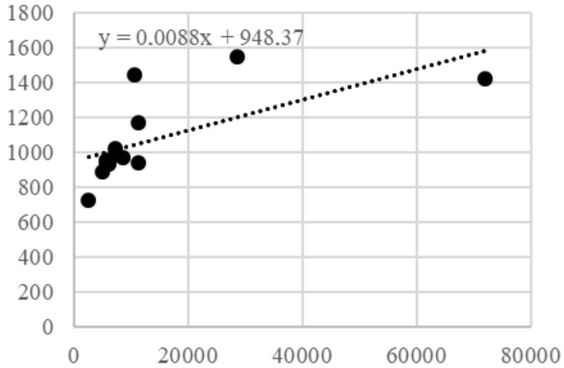

Fig. 6. Dependence of the volume of transport services per capita on the volume of industrial production

As shown in Figure 7, the dependence of the volume of transport services per capita on the volume of retail trade turnover per capita is strong (correlation coefficient is 0.795605). The higher the retail turnover, the greater the volume of transport services.

The dependence of the volume of transport services per capita on the volume of exports per capita is moderately positive (correlation coefficient is 0.61924), there is a direct impact of foreign trade on the volume of transportation (Figure 8).

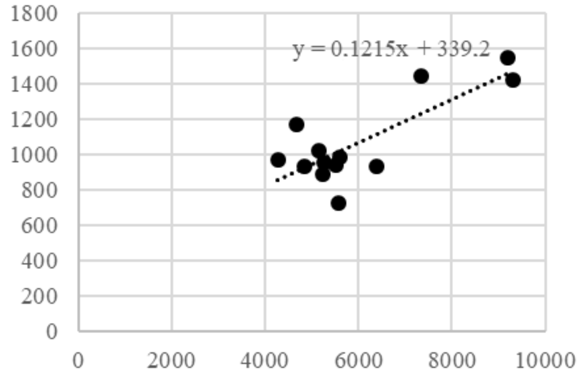

Fig. 7. Dependence of the volume of transport services per capita on the volume of retail trade turnover per capita

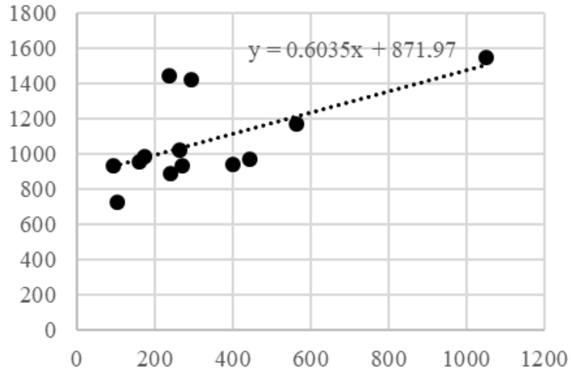

Fig. 8. Dependence of the volume of transport services per capita on exports per capita

From Figure 9, it can be seen that the dependence of the volume of transport services per capita on the volume of imports per capita is moderately positive (correlation coefficient is 0.662933), import directly affects the volume of transportation.

In Figure 10, there is a weak negative correlation between the volume of transport services per capita and the physical volume of freight transportation by rail (correlation coefficient is minus 0.07812).

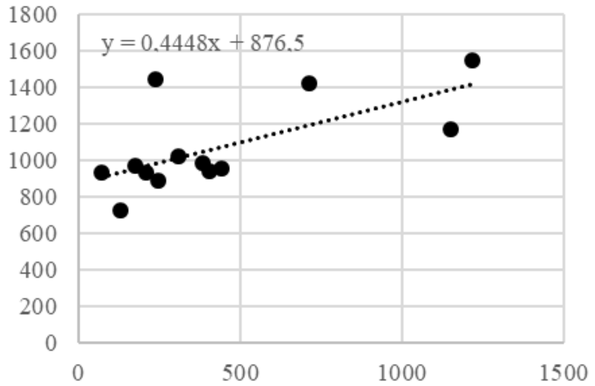

Fig. 9. Dependence of the volume of transport services per capita on the volume of imports per capita

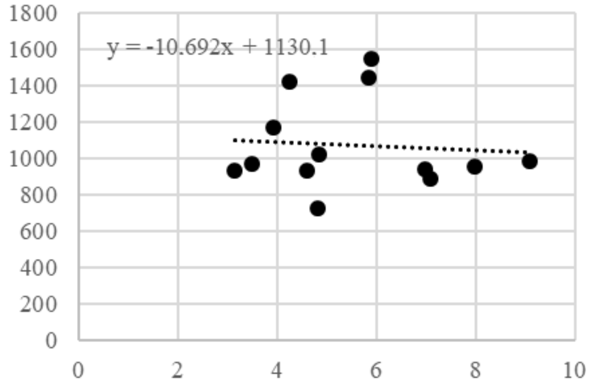

Fig. 10. Dependence of the volume of transport services per capita on the volume of cargo transportation by rail

Figure 11 shows a weak positive dependence of the volume of transport services per capita on the physical volume of cargo transportation by rail (correlation coefficient is 0.124004).

In Figure 12, there is a weak positive correlation between the volume of transport services per capita and the density of the railway network (correlation coefficient is 0.19887). There is no dependency.

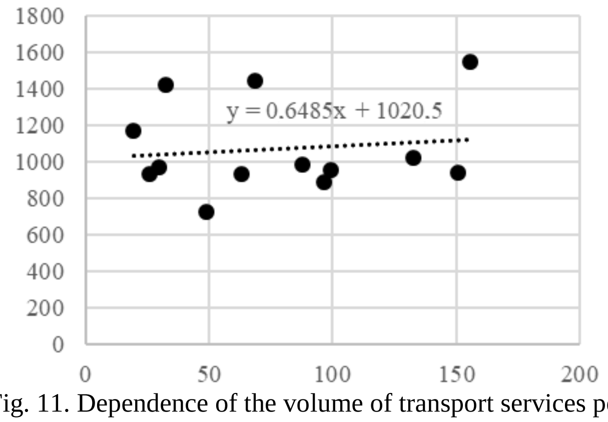

Fig. 11. Dependence of the volume of transport services per capita on the volume of cargo transportation by road

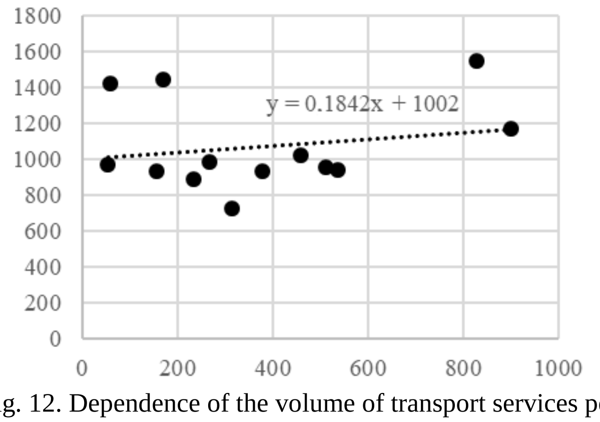

Fig. 12. Dependence of the volume of transport services per capita on the density of the railway network

Figure 13 shows a weak negative correlation between the volume of transport services per capita and the density of roads (correlation coefficient is minus 0.21887); there is no dependence.

The dependence is not observed in Figures 14 and 15 with correlation coefficients of 0.46175 and minus 0.0871, respectively.

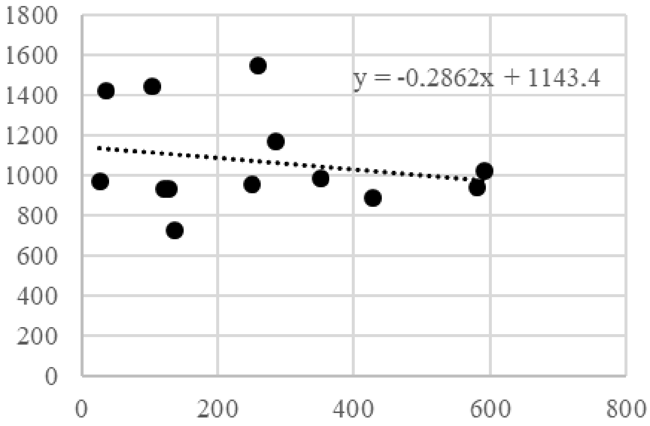

Fig. 13. Dependence of the volume of transport services per capita on the density of roads

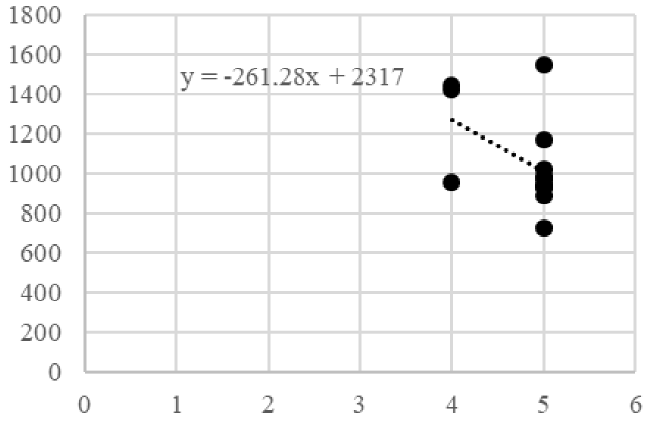

Fig. 14. Dependence of the volume of transport services per capita on the availability of transport corridors

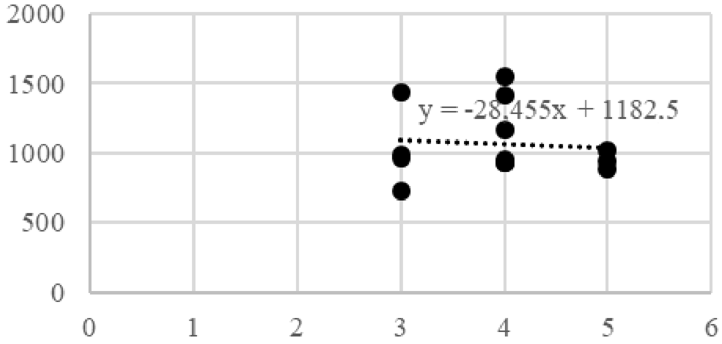

Fig. 15. Dependence of the volume of transport services per capita on the climatic zone

The main functions of the dependence on factors influencing the location of logistics centers are presented in Table 1.

Table 1. The main functions of the dependence of factors influencing the location of logistics terminals in 2022

Equation	Multiple correlation coefficient, R	Multiple correlation coefficient, R^2	F-test significance	Durbin-Watson statistic, DW
Social-economic factors				
$ChN = ChN_0 + ChR - ChU$	1	1	7.37E-129	0.39
$D = -221657 + 1.44799*VRP - 0.78514*OP + 2908*r_1 - 985.258*r_3 + 275.1267*r_4$	0.932	0.868	0.008041	2.71
$VRP = 23799 + 0.5365*OP - 6.8496*I_{inv} + 0.4717*OT + 0.7144*ChNZE + 7.522*OF + 13694*r_1 + 777.49*r_2 + 249.09*r_3$	0.999	0.998	0.000461	2.52
$OP = -307327 + 5.3903*OF + 16.6843*ChNZE - 815574*r_2 - 105922*r_3 - 966446*r_4$	0.848	0.7196	0.072447	1.46
$OT = 899{,}7929 + 0{,}2225*D + 0{,}0275*OP + 1{,}9653*E + 2{,}014*Pl_{avt} + 1473{,}689*r_2 + 598{,}02*r_3 + 347{,}35*r_4$	0.8796	0.774	0.202727	2.09
$E = 991.41 + 0.0064*OP + 217.217*TK + 0.4858*Pl_{jd} - 98.3*r_1 - 160.199*r_2 + 43.54*r_3$	0.781	0.609	0.322535	2.31
$I = 398.903 - 0.0549*OT + 0.01295*VRP - 0.034*ChN + 1.188*E + 118.21*KZ - 411.997*r_1 + 303.85*r_2 + 127.08*r_3$	0.885	0.784	0.340326	1.55

$V_{tr.usl} = 102.0186 + 0.046 * D - 0.00185 * VRP + 0.3606 * Pl_{avt} + 0.435 * E - 4.711 * r_1 + 192.84 * r_3 + 25.305 * r_4$	0.941	0.886	0.058044	2.07
$OF = -2827.83 + 0.4411 * D + 0.07396 * OP + 1624.94 * r_2 + 3148.36 * r_3 + 2097.02 * r_4$	0.893	0.798	0.028323	2.04
$I_{inv} = -6578.35 + 1.108 * VRP - 0.588 * OP - 0.012 * OT - 269.883 * r_1 + 1104.14 * r_3 + 1010.87 * r_4$	0.969	0.940	0.004052	1.86
$ChNZE = 73.3146 + 0.40597 * ChN + 33.604 * KZ + 10.328 * r_1 - 10.7523 * r_3 - 34.7827 * r_4$	0.995	0.991	2.975E-06	1.93
Transportation indices				
$V_{per.jd} = 17.957 - 0.005 * Pl_{avt} - 0.0003 * Pl_{jd} - 0.00029 * I + 0.0005 * OF + 0.063 * U_{per.avt} - 0.0014 * OT - 1.953 * TK + 2.835 * r_1 - 2370 * r_3$	0.980	0.960	0.057949	2.61
$V_{per.avt} = -408916 + 0.039 * ChN - 1.626 * TK + 0.023 * OT + 0.097 * Pl_{avt} - 0.002 * D - 0.006 * OF - 38.989 * r_2 - 10.0205 * r_3 - 15.0929 * r_4$	0.986	0.972	0.033467	2.04

The variables and symbols of the equations of the statistical model are presented in Table 2.

Table 2. Variables and Symbols of the Statistical Model

№	Designation	Variable	Unit measurements
Exogenous (independent) variables			
1	KZ	Belonging to the climatic zone	score
2	TK	Availability of transport corridors on the territory of the subject	score
3	Pl_{avt}	Provision of the subject with railway tracks	km/10000 m2
4	Pl_{jd}	Availability of paved roads	km/1000 m2
5	ChN	Population	thousand people
6	ChR	The number of people born per 1000 people per year	thousand people
7	ChU	The number of deaths per 1000 people per year	thousand people
Endogenous (dependent) variables			
8	D	Average per capita income of the population	soum/person
9	VRP	Volume of gross regional product per capita	soum/person
10	OP	Volume of industrial production per capita	soum/person
11	OT	Volume of retail trade turnover per capita	soum/person
12	E	The volume of export products per capita in the region	USD/person
13	I	The volume of imported products per capita in the region	USD/person
14	$V_{tr.usl}$	Volume of transport services per capita	soum/person
15	OF	Fixed assets per capita	soum/person
16	I_{inv}	Investment in fixed assets per capita	soum/person
17	ChNZE	Population employed in the economy	thousand people
18	$V_{per.jd.}$	The volume of goods transported by rail per capita	t/person
19	$V_{per.avto}$	The volume of goods transported by road per capita	t/person
Auxiliary variables			
	r_1	A dummy variable denoting the subject's membership in the first cluster	-
	r_2	A dummy variable denoting the subject's membership in the second cluster	-
	r_3	A dummy variable denoting the subject's membership in the third cluster	-
	r_4	Dummy variable denoting the subject's belonging to the fourth cluster	-

To check the adequacy and accuracy of the developed model, calculations were performed for all model equations: the multiple correlation coefficient (R), characterizing the closeness of the relationship between the dependent variable and the set of indicators that determine it; coefficient of determination (R^2), characterizing the reliability of the statistical model; F-test significance for testing the hypothesis on the dependence of the variables and the predicted indicator; the Durbin-Watson (DW) statistic to assess the presence of autocorrelation.

All calculations were performed in Excel; when the initial data changes, the developed template can be used to assess the situation of previous periods and to forecast for the future periods.
Based on the results of the calculations, we can draw conclusions confirming the viability of the model:
- since the coefficient for all equations of the model is greater than 0.8, and for the majority is more than 0.9, the model reveals at least 80-90% of the variations of dependent variables;
- 95% reliability was established - the significance of the coefficients of the regression equation and its free term at the 5% level;
- the level of the standard error of estimating dependent variables according to the equations does not exceed 5% of the average value of the variable;
- the remainders of the regression equation correspond to the normal distribution law.

4. CONCLUSIONS

Thus, the equations of the developed statistical model made it possible to establish the dependence of the volume of transport services in the regions of the Republic of Uzbekistan on the income of the population, the gross regional product, the density of roads, the volume of exports and belonging to clusters. For the second cluster (Samarkand region), there is no such dependence.
The dependence of rail transport volumes on the density of roads and railway coverage, imports, the cost of fixed assets per capita, road transport volumes, retail trade volumes and the availability of transport corridors was statistically confirmed. At that, the indicator increases due to belonging to the first cluster (Khorezm region) and decreases due to belonging to the third cluster (Republic of Karakalpakstan, Andijan, Bukhara, Jizzakh, Kashkadarya, Navoi, Surkhandarya, Syrdarya, Tashkent, Fergana regions).
In the course of the study, it was found that the volume of cargo transportation by road is determined by the cost of fixed assets per capita, the number of people employed in the economy, trade volumes, per capita income of the population, the presence of transport corridors and the density of roads. The subject's belonging to the second, third, and fourth clusters (Namangan region) reduces the values of these indicators. This is explained by the high value (in comparison with road transport) of the length of railways, and the transportation is more profitable.
It is appropriate to study the interaction of the proposed factors in dynamics using the simulation method. This is justified by the fact that it is difficult to express the assessment of the development of the system of factors for the location of cargo objects by analytical dependencies.
The use of the dependencies stated in the process of research in the development of a simulation model will allow us to establish the influence of factors on the formation of demand for logistics infrastructure facilities in dynamics and obtain predictive values of indicators to assess the prospects for the development of cargo facilities in the regions of the Republic of Uzbekistan.

5. REFERENCES

1. Bureeva, N.N., (2007). *Multivariate statistical analysis using “STATISTICA”*, Educational material, pp. 112, Russia (Nizhny Novgorod).
2. Fayzibaev, S., Mukhamedova, Z., Ergasheva, Z., (2022). *Improving the Design Concepts of Equipment for the Assembly Platform of a Rail Service Car Considering Reliability Rates and Real State*, IP Conference Proceedings, 2432, 030052.
3. Ivanova, T.A., Dorogina, K. Yu., Popova, I.N., Druzhinin, Yu.D., (2012). *Methods for studying socio-economic and demographic processes*, FGBOU VPO “Magnitogorsk State. Tech. University named after G.I. Nosov”, pp. 155, Russia (Magnitogorsk).
4. Makhkamov, N. Y., Ibragimova, G. R., & Ismatullaev, A. F. (2020). The rational connection coefficient calculation with different train structures. In IOP Conference Series: Materials Science and Engineering (Vol. 918, No. 1, p. 012052). IOP Publishing. https://doi.org/10.1088/1757-899X/918/1/012052.
5. Mukhamedova, Z.G. (2016). *Modelling of fluctuations in the main bearing frame of railcar*, Int. J. of Mod. Manufact. Technol., VIII (2), 48-53.
6. Mukhamedova, Z., (2020). *Analysis and assessment of power efficiency of special self – propelled railway rolling stock*, Journal of Advanced Research in Dynamical and Control Systems, 12(2), 2808–2814.
7. Mukhamedova Z.G., Ibragimova G.R., (2022). *On the issue of forming the cargo infrastructure of the Republic of Uzbekistan*, Transsib News, I (49), 59 - 66.

8. Saburov, M., Butunov, D.B., Khudayberganov, S., Boltaev, S.T., Akhmedova, M., & Musaev, M. (2022). *Determination of the optimal requirement of the number of freight wagons*, In AIP Conference Proceedings (Vol. 2432, No. 1, p. 030091). AIP Publishing LLC. https://doi.org/10.1063/5.0090343.
9. Sai, V.M., (2015). *Integrated coefficient of project efficiency in the interaction of Russian Railways with regional economic entities*, Transport of the Urals. Vol. (No 4), pp. 9.
10. Sai, V.M., Sizyi, S.V., Fomin, V.K., (2010). *Integral assessment of enterprises*, Economics of railways. Vol. (No. 1), pp. 18.
11. Suyunbayev, Sh., Khusenov, U., Khudayberganov, S., Jumayev, Sh., Kayumov, Sh., (2023). *Improving use of shunting locomotives based on changes in infrastructure of railway station*, E3S Web of Conf., Volume 365, 05011, DOI: https://doi.org/10.1051/e3sconf/202336505011.

Received: February 21, 2023 / Accepted: April 05, 2023 / Paper available online: June 20, 2023 © International Journal of Manufacturing Economics and Management.