

Mathematical model for prediction of cargo flow during the construction of the railway line Uzbekistan - Kyrgyzstan – China

Khasan Umarov^{1*}, *Yulduz Tursinaliyeva*¹, *Islom Khurramov*¹, and *Saulet Shayakhmetov*²

¹Tashkent State Transport University, Tashkent, Uzbekistan

²Satbayev University, Almaty, Kazakhstan

Abstract. The purpose of this work is to study a mathematical model for predicting freight traffic during the construction of a new railway line Uzbekistan - Kyrgyzstan - China, taking into account the considered economic factors. The proposed algorithm for constructing a predictive mathematical model of freight traffic during the construction of a new Uzbekistan-Kyrgyzstan-China railway line will make it possible to choose technical solutions that determine the operation of the Uzbekistan railway in the future, as well as the best options for strengthening its capacity in the future.

1 Introduction

Justification of the technical and economic new railway line Uzbekistan - Kyrgyzstan - China, one of the most important tasks is to determine the transit freight traffic between the countries of China, Central and South Asia. The economic uncertainty in the world at the moment and the uncertainty of the timing of the construction of a new railway line Uzbekistan-Kyrgyzstan-China do not allow determining the size of the transit traffic of China, Central and South Asia [1-4].

The study of modern world experience makes it possible to substantiate in the current conditions the prospects for the long-term development of investment projects implemented in the field of railway transport, in order to ensure the continuity and efficiency, as well as the profitability of their work, the maximum reliability of the predictive and planned initial information on the size of freight traffic for the future is necessary [5-8].

The reliability of forecasts of information on the size of cargo traffic for the future is considered the main task for a qualitative analysis and justification of investment projects used in modern conditions, prevents the possibility of uncertainties in the future, and minimizes possible risks in the implementation of projects. This has a great influence on the choice of the design capacity of the projected railway, as well as on operational performance, the prospective development of the project and the degree of validity of measures to increase the capacity of the roads [9-13].

*Corresponding author: janobhuk@mail.ru

2 Objects and methods of research

According to the sources, there are two groups of forecasting methods [14-16]: formalized and intuitive.

Intuitive methods can be applied to determine the size of the transit traffic of China, Central and South Asia during a period of severe economic uncertainty in the world. Because today it is very difficult to determine the factors that influence the determination of the size of the transit traffic in China, Central and South Asia. In such cases, the results of the transit cargo flow are accepted by individual and collective expert opinions.

The first method, correlation-regression analysis, involves the formation of a mathematical model that formalizes the causal mechanism of the indicator behavior using the methods of correlation and regression analysis or other statistical methods for studying factor dependencies [17, 18]. The first method is advisable to apply under stable economic conditions and short-term forecasting of transport traffic.

The second method of predictive extrapolation is based on the ability to predict the future by studying the patterns of change in the time series of an indicator without the influence of any factors using linear growth indicators. This method is used when there is insufficient statistical information about changes in influencing factors in the current market conditions.

Because economic uncertainty in the world is currently affecting the unstable market relations between China, Central and South Asia, it is still very difficult to make a reliable forecast of transit traffic [17, 18]. The volume of transit cargo flows between China, Central and South Asia should be predicted taking into account the influence of the most important economic factors. Thus, the period from 2023 to 2027 is determined by the method of correlation and regression analysis based on the influence of economic factors on the forecast of transit traffic between China, Central and South Asia. The forecast of transit cargo traffic between China, Central and South Asia in the period from 2028 to 2035 is determined by the predictive extrapolation method.

In order to build a model for predicting freight traffic on the new Uzbekistan-Kyrgyzstan-China railway line, we will divide it into two directions:

- Cargo flow of China and South Korea to Uzbekistan;
- Cargo flow of China with Tajikistan, Afghanistan, Turkmenistan and Iran.

To build a predictive mathematical model of the "behavior" of the freight traffic of the railway direction, it is necessary to use a systematic algorithm.

At the first stage, an analysis of the object is carried out, on the basis of which factors are selected that characterize the main content of the predicted values of the freight traffic of the railway direction, as well as the preparation of an information base.

Statistical data on the transit traffic on the new railway line Uzbekistan-Kyrgyzstan-China from 2012 to 2022 (Fig. 1).

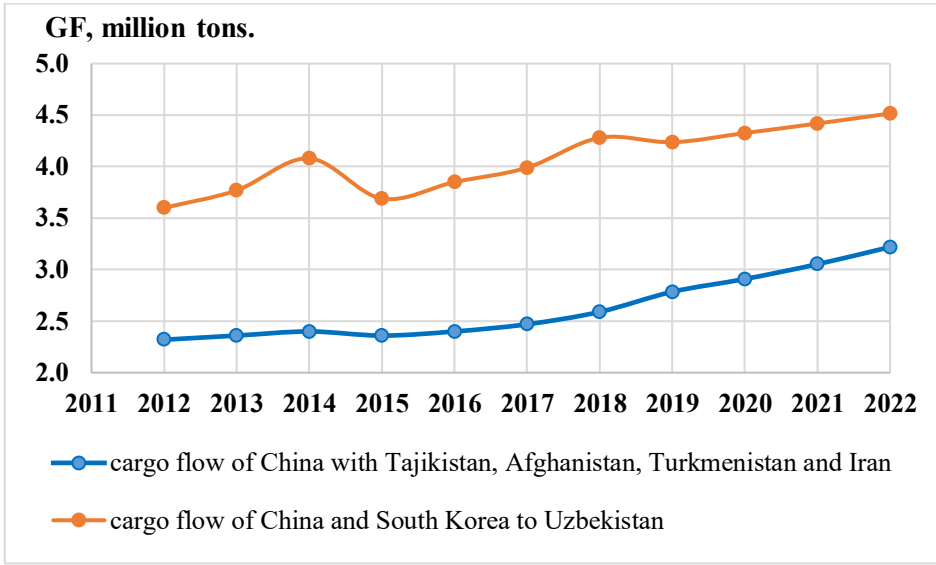


Fig. 1. Actual cargo flow of the studied landfill for the period from 2012 to 2022

Economic factors influencing the gross domestic product (GDP) of the countries under consideration ($X_C, X_{S,K}, X_U, X_T, X_A, X_{Tr}, X_I$) are proposed as the main factors for forecasting freight traffic on the new railway line Uzbekistan-Kyrgyzstan-China. According to the GDP statistics [19] of the countries under consideration for the period from 2012 to 2022 are summarized (Table 1).

Table 1. GDP statistics for the countries under consideration for the period from 2012 to 2022, USD billion

Year	Factor X_C GDP China	Factor $X_{S,K}$ of GDP South Korea	Factor X_U of GDP Uzbekistan	Factor X_T GDP Tajikistan	Factor X_A GDP Afghanistan	X_{Tr} factor of GDP Turkmenistan	Factor X_I GDP Iran
2012	8550	1223	51.6	7.3	19.9	33.5	483.8
2013	8939	1305	55.2	8.5	20.7	40.6	411.9
2014	10360	1452	63.1	9.2	21.7	47.5	402.7
2015	11075	1561	68.2	9.3	21.8	52.2	393.5
2016	11 218	1 411.2	66.5	6.9	18.9	36.2	376.8
2017	12 015	1 538.0	47.9	7.3	20.9	37.9	431.9
2018	13 608	1 619.4	50.5	7.5	19.4	40.8	452.3
2019	14732	1646.7	57.9	8.1	18.9	45.2	583.7
2020	14 867	1 638.3	59.9	8	20.1	45.6	835.4
2021	17745	1811	69.2	8.7	14.9	62.2	838.5
2022	18277	1817	80.4	11.3	12.6	66.0	863.6

The assessment of the influence of factors is carried out because of a correlation-regression analysis of the freight traffic on the new railway line Uzbekistan - Kyrgyzstan - China, which characterizes the close relationship between the transit traffic and the factors affecting it. In order to determine the presence of a correlation between the statistical data of influencing factors and the indicator of transit cargo traffic of the investigated new railway line Uzbekistan-Kyrgyzstan-China, we calculate the correlation coefficient [20]:

$$r_{xr} = \frac{\sum (x_i - \bar{x}) \cdot (gf_i - \overline{gf})}{\sqrt{\sum (x_i - \bar{x}) \cdot \sum (gf_i - \overline{gf})}} \quad (1)$$

Where x_i are the values taken in the X sample; gf – values accepted in the sample GF ; \bar{x} – average value for X ; \overline{gf} – the average value for GF .

In order to establish the degree of dependence between the considered statistical data of the factors $X_C, X_{S,K}, X_U, X_T, X_A, X_{Tr}$ and X_I , and indicators of the freight traffic of railway transport, correlation coefficients were calculated.

The obtained correlation coefficients have positive [20, 21] values, mostly close to unity, which indicates a close relationship between the traffic and the factors affecting it.

At the second stage of assessing the influence of factors, a correlation-regression analysis of the freight traffic of the railway direction is carried out, which characterizes a strong close relationship between the freight traffic and the factors influencing it, as well as the compilation of matrices for constructing a multiple regression equation. The multiple linear regression equation is as follows [20, 21].

$$GF = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k; \quad (2)$$

where $\alpha_1, \dots, \alpha_k$ – linear regression coefficients; x_1, x_2, \dots, x_k – independent variables (economic factors); GF – dependent variable (value of freight traffic).

Also, the compilation of matrices for constructing a multiple regression equation. We compile a matrix of the studied factors - matrix X and a matrix of cargo traffic data - matrix GF .

It is necessary to define matrices $|X^T \cdot X|$. In order for the regression equation to have a solution, the determinant of the matrix must not be equal to zero, i.e. the matrix must be non-singular [20, 21].

We determine in a matrix way the unknown coefficients of the equation by the formula:

$$A = (X^T \cdot X)^{-1} \cdot X^T \cdot GF \quad (3)$$

We determine in a matrix way the unknown coefficients of the equation by the formula:

Where A – column vector of coefficients of the regression equation; X^T – transposed matrix X ; X – dimension matrix n – lines and $(k+1)$ – columns of known influencers.

At the third stage, dependencies are checked for the adequacy and significance of the regression coefficients (F – criterion) between the traffic and the factors affecting it. The construction of the regression equation is reduced to estimating its coefficients. The least squares method is used to estimate the regression coefficients.

$$F_{cal} = \frac{\frac{1}{k+1} \cdot Q_R}{\frac{1}{n-k-1} \cdot Q} > F_{crit} \quad (4)$$

$$Q_R = \sum_{i=1}^n (GF)^2 \quad (5)$$

$$Q = \sum_{i=1}^n (GF - \overline{GF})^2 \quad (6)$$

where F_{crit} – analysis of variance criterion (F – criterion); k – number of factors ; Q_R – the sum of squared deviations from zero, calculated by the formula (5); n – number of observations equal to 11; Q – the sum of the squared deviations [20, 21] of the observation results, calculated by the formula(6).

The fourth stage is the calculation of forecast traffic data (until 2027) according to the found multiple regression equation.

At the fifth stage, the development trend of cargo traffic for the period from 2023 to 2027 is determined. To describe the development trend of the phenomenon, linear or curve models are widely used, which are various functions of time $y=f(t)$.

To solve this problem, you first need to choose the type of function. The most commonly used functions are [21]:

- polynomial of the first degree (linear): $GF(t) = a_0 + a_1t$;
- polynomial of the second degree: $GF(t) = a_0 + a_1t + a_2t^2$;
- polynomial of the third degree: $GF(t) = a_0 + a_1t + a_2t^2 + a_3t^3$.

Model quality can be checked in the same way as for a paired regression model, checking the statistical significance of the coefficients and the overall quality using the coefficient of determination - R^2 . The value of the coefficient of determination varies from 0 to 1 and shows to what extent the dynamics of the resultant attribute is described by the dynamics of the factorial one. Determination coefficient:

$$R^2 = \frac{\sum(\widehat{GF}_{C-U} - \overline{GF}_{C-U})^2}{\sum(GF_{C-U} - \overline{GF}_{C-U})^2} \quad (7)$$

To assess the accuracy of the model, we calculate the average relative approximation error [22, 23]:

$$\delta = \frac{1}{n} \sum_{i=1}^n \frac{|e_i|}{GF_{C-U}} \cdot 100\% \quad (8)$$

The actual values of the effective indicator differ from the values calculated according to the model equation by the value $e_i = GF_{C-U} - \widehat{GF}_{C-U}$. This value in each observation is the absolute approximation error. However, these quantities are incomparable among themselves, since they depend on the units of measurement and the scale of the quantities GF_{C-U} .

At this stage, the simulation results are evaluated and the statistical correctness of the constructed mathematical model is checked using the following coefficients: Student's criterion, model accuracy estimates, standard error of the estimate, multiple correlation coefficient.

$$S = \sqrt{\frac{\sum e_i^2}{n-1}}; \quad (9)$$

where S – trend standard error of the trend; n – time series observations. Let's check the statistical significance of the coefficient α_0 .

$$S_{\alpha_0}^2 = \frac{S^2}{\sum(t_i - \bar{t})^2}; \quad (10)$$

According to statistics

$$t_{\alpha_0} = \frac{|a|}{S_a}; \quad (11)$$

The last stage is the verification of mathematical models for forecasting cargo traffic using extrapolation from time series using the least squares method.

3 Results and their discussion

The correlation coefficient of the China-South Korea freight traffic indicator between Uzbekistan – GF_{C-U} with the effective factors X_U is $r_{xg}=0.822$, with X_C is $r_{xr}=0.969$, with $X_{S,K}$ is $r_{xg}=0.957$, which accordingly characterizes the strong closeness of the relationship between the studied indicators on the scale Chaddock. Moreover, the result of the obtained correlation coefficients is represented by the transit traffic of China between Tajikistan, Afghanistan, Turkmenistan and Iran – $GF_{C-TATrI}$ with effective factors X_C is $r_{gr}=0.961$, with X_T it is $r_{gr}=0.930$, with X_A it is $r_{gr}=0.809$, with X_{Tr} it is $r_{xg}=0.933$, with X_I is $r_{xg}=0.812$, respectively. In the course of the study, it was found that the significance of the regression coefficients obtained from the mathematical model was confirmed when determining the transit traffic of the considered new railway line Uzbekistan-Kyrgyzstan-China.

From the equation of multiple regression of the transit traffic of China and South Korea with Uzbekistan, the following result was obtained:

$$GF_{C-U} = 0.00035 \cdot X_U - 0.0000819 \cdot X_C - 0.0032893 \cdot X_{S,K}$$

China between Tajikistan, Afghanistan, Turkmenistan and Iran:

$$GF_{C-TATrI} = 0.000174 \cdot X_C + 0.19318 \cdot X_T - 0.077 \cdot X_A - 0.0269 \cdot X_{Tr} - 0.000012 \cdot X_I$$

At a significance level of 0.05 [20, 21], we will check the significance of the regression equation, the null hypothesis, using the F-criterion. According to expression (4). According to the table F - distributions [20, 21] for a significance level of 0.05 and degrees of freedom 5 and 7, respectively, we find the critical value: $F_{crit} (0.05; 5; 7) = 4.53$. Hence $F_{cal} > F_{crit}$; therefore, the hypothesis H_0 is rejected.

According to the forecasts of the International Monetary Fund (IMF), the total transit cargo flow between China, South Korea and Uzbekistan will be received in the period from 2023 to 2027. (Table 2). It should be noted that according to IMF forecasts, from 2023 to 2027, GDP would grow gradually, which will affect the total transit traffic between China and Central and South Asia.

An error of less than 7%-10% indicates a good fit of the model to the original data (good accuracy). If the error is more than 12-15%, you should think about choosing a different type of model equation. The calculation of the accuracy of the constructed model averages $\delta=2.54\%$, which indicates sufficient accuracy of the constructed model.

Table 2. The results of the calculation of the forecast cargo flow between China, South Korea and Uzbekistan for the period from 2023 to 2027, million tons.

Year	Factor X_C GDP China	Factor $X_{S,K}$ of GDP South Korea	Factor X_U of GDP Uzbekistan	Factor X_T GDP Tajikistan	Factor X_A GDP Afghanistan	X_{Tr} factor of GDP Turkmenistan	Factor X_I GDP Iran
2023	19410	1872	84	12	14	70	870
2024	20614	1928	89	13	14	74	876
2025	21892	1985	93	14	15	79	882
2026	23249	2045	98	15	15	84	888
2027	24690	2106	103	16	15	89	895

The statistic t_{α_0} is greater than the tabular value of the Student statistic. According to expression (11) $t(16; 0.05)=4.33$. Consequently, the coefficient α_0 is statistically significant with a probability of 95% differs from zero, which confirms the presence of the dependence of the indicator GF_{C-U} on time.

The study of various fitting curves showed a strong relationship between polynomials of the second and third degrees, which better describe the dependence of the transit traffic between China and Central and South Asia for the period from 2023 to 2027 (Fig. 2 and 3). In the course of studying the long-term forecast of transit traffic between China and Central and South Asia (until 2035), the forecast took the form of a polynomial of the 3rd degree in the "optimistic scenario", the form of a polynomial of the 2nd degree in the "pessimistic scenario", and their average values considered as a "base scenario".

Let us analyze the obtained statistical results. In addition, it should be noted that the economic crisis of 2015 had a lesser impact on the transit traffic between China and Central and South Asia. An analysis is needed, will they affect the occurrence of the 2020 pandemic (Covid-2019), the volume of China's transit cargo flow between Central and South Asia. It is necessary to analyze the results obtained every 5 years, taking into account the long-term forecasting of transit traffic between China, Central and South Asia. If the difference in results exceeds 15%, transit traffic must be corrected using the new statistics.

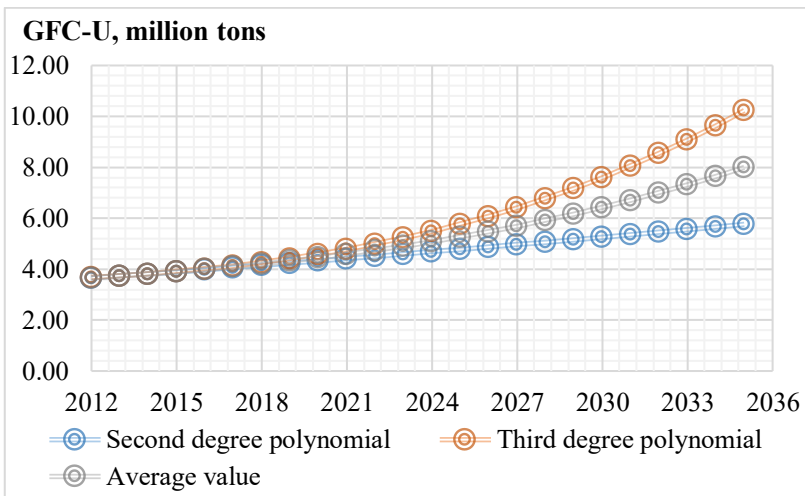


Fig. 2. Forecast of cargo traffic between China, South Korea and Uzbekistan for the period from 2023 to 2035

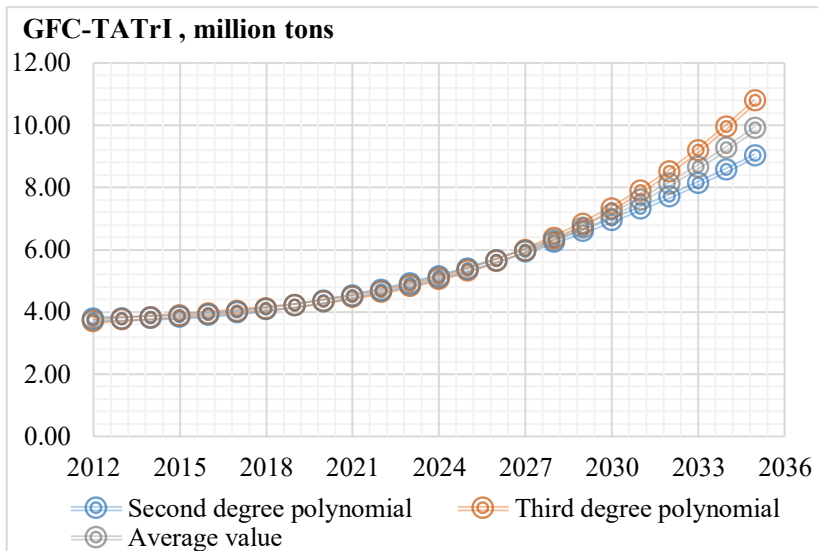


Fig. 3. Forecast of cargo traffic between China and Tajikistan, Turkmenistan, Afghanistan and Iran for the period from 2023 to 2035

4 Conclusions

1. An algorithm for constructing a predictive mathematical model of the "behavior" of the freight traffic of the railway direction has been developed.

2. With the help of statistical and correlation-regression analysis, economic factors were identified and proposed that affect the volume of transit cargo flows of the new railway line Uzbekistan-Kyrgyzstan-China.

3. With the help of the developed mathematical model, the following results of the forecast of the transit cargo flow of the new railway line Uzbekistan-Kyrgyzstan-China for the period from 2023 to 2035 were obtained:

- in 2025, the volume of transit cargo traffic between China, South Korea and Uzbekistan by rail will amount to 5.27 million tons, in 2030 - 6.43 million tons, in 2035 - 7.99 million tons;

- transit cargo traffic between China and Tajikistan, Turkmenistan, Afghanistan and Iran for 2025 is 5.35 million tons, for 2030 - 7.14 million tons, for 2035 - 9.91 million tons.

4. Using the developed mathematical model, the transit cargo flow of the new Uzbekistan-Kyrgyzstan-China railway line is 10.62 million tons for 2025, 13.57 million tons for 2030, and 17.90 million tons for 2035.

References

1. Umarov Kh.K., Svintsov E.S. Mathematical model for the traffic forecasting China and South Korea between Central and South Asia /VESTNIK RGUPS. Scientific and Technical Journal. – 2017 – Issu. 2(66). – pp. 69-75.
2. Lesov K., Kenjaliev M., Mavlanov A. & Tadjibaev Sh. Stability of the embankment of fine sand reinforced with geosynthetic materials. E3S Web of Conferences, 2021, no. 264, pp. 02011, <https://doi.org/10.1051/e3sconf/202126402011>.

3. Lesov K.S. & Kenjaliyev M.K. Organizational and Technological Parameters during the Construction of the Bukhara-Misken Railway Line. AIP Conference Proceedings, 2022, no. 2432, pp. 030026. DOI: 10.1063/5.0089621.
4. Muzaffarova M. & Mirakhmedov M. Differences and commonalities impregnation of dry and wet sand. Transport Problems, 2014, vol. 9, no. 3, pp. 91-97.
5. Muzaffarova M. & Mirakhmedov M. Prospect's fixation drift sands physicochemical method. Transport Problems, 2016, vol. 11, no. 3, pp. 143-152, <https://doi.org/10.20858/tp.2016.11.3.14>.
6. Lesov K.S., Mavlanov A.Kh. & Kenjaliyev M.K. Load-Bearing Capacity of Pile Foundations of Contact Network Supports in the Sandy Soil Subgrade. AIP Conference Proceedings, 2022, no. 2432, pp. 030027, DOI: 10.1063/5.0089622.
7. Shaumarov S., Kandakhorov S. & Umarov K. Development of the optimal composition of aerated concrete materials based on industrial waste. AIP Conference Proceedings, 2022, no. 2432, pp. 030087, DOI: 10.1063/5.0089875.
8. Lesov K., Ortikboyev M. & Yuldasheva N. Design of Rockfall Protection Structures. ICNTET, 2022, no. 1210, pp. 4222209.
9. Ablyazov V.K. Forecasting of port cargo turnover in conditions of risk and uncertainty: dis. candidate of Economic Sciences. 2013. – 185 pp.
10. Khalfin G.-A., Umarov K., Purtseladze I. & Botirov O. The Work of Intermediate Rail Fasteners on Difficult Sections of Railways. ICNTET, 2022, no. 1234, pp. 4222233.
11. Kakharov Z. V., Eshonov F. F., Kozlov I. S. Determination of the values of the energy constants of materials during crushing of solids. News of the St. Petersburg University of Railway Transport. 2019. Vol. 16. No. 3. – pp. 499-504.
12. Kakharov Z.V. Railway construction for high-speed roads. Universum: technical sciences: electron. Scientific. Journal. 2022. 5(98).
13. Djabbarov S., Kakharov Z., Kodirov N. Device of road boards with compacting layers with rollers. AIP Conference Proceedings. AIP Publishing LLC, 2022. – Vol. 2432. № 1. p. 030036. <https://doi.org/10.1063/5.0089679>
14. Anderson T. Statistical analysis of time series. M.: Mir, 1976. 757 p.
15. Buslenko N.P. Modeling of complex systems. M.: Nauka, 1978. 400 pp.
16. Bykov Yu.A. Basic system-technical principles of railway design. Improving the theory and practice of designing, reconstruction and operation of railways. Intercollegiate Collection of scientific works Khabarovsk, 1994, pp. 17-23.
17. Kh. Umarov. Mathematical model for forecasting freight flows between Ferghana valley and other regions of Uzbekistan. Philosophical Readings XIII 4 (2021): pp.1318-1328.
18. Umarov X., Botirov O. The role of construction of the angren-pap railway line in the plans of international transport and economic relations. Universum: technical science. 2021. № 6-5 (87).
19. Kremer N. S. Probability theory and mathematical statistics: textbook for university students studying in economic specialties. 3rd ed., reprint. and additional M.: UNITY-DANA, 2007. - 551 pp.
20. E. Efimova.N forecasting freight traffic volumes using econometrics techniques // Economics of Railways, 2005. No. 6. pp. 31-43.
21. E. Chetverkina.M. Statistical methods of forecasting. Second edition, reprint. and additional. M.: Statistics, 1977. p.200.