ACTA TTPU

III. ENGINEERING SCIENCE

APPLICATION OF THE LINEAR REGRESSION METHOD TO DETERMINE THE EFFECTIVE ORGANIZATION OF THE TRANSPORTATION

R. Samatov[1], A. Hoshimov[2], I. Kambarov[2], J. Inoyatkhodjaev[2]

[1]Tashkent Institute of Design, Construction and Maintenance of Automobile Roads
[2]Turin Polytechnic University in Tashkent
E-mail: samrustar@mail.ru, j.inoyatkhodjaev@polito.uz, ikrom.kambarov@polito.uz, a.hoshimov@polito.uz

Abstract

Within the operations of a transportation company, there is uncertainty about how much product will be sold, manufactured and shipped. Even with improved forecasting methods it is not possible know precisely how much demand there will be for a company's products in the future. This means that it is difficult to know how much transportation expense, time, distance the firm will require, or the impact of the transportation elements would have on total accuracy of the transportation. Therefore it requires to investigate the framework to determine effectively organization of the transportation systems. This paper presents analysis and application of the linear multifactorial regression as a method to determine the profit of the transportation system, depending on the performance of its operation.

Keywords: transport problems, methods of organizing transportation, performance, level of service, indicators, linear regression

Introduction

The implementation of supply chain network for raw materials, semi-finished products and distribution of finished products within the framework of the logistics system requires solving a complex transportation problems. [1], [2], [3]. In a supply chain network, companies cooperate with many suppliers and customers, and interact through a variety of information and material flows and has to deal with issue related to delivery, to achieve balance between supply and demand [4]. In general supply chain network breaks down into a series of consecutive specific individual stages, which are unrelated activities to each other's and performed by different departments. The optimization of such complex and space time systems requires a comprehensive approach and need to quantitatively describe.

Nowadays, the issues of increasing the level of transport customer service, which in market conditions are closely connected with the issue of service and quality of services provided, are becoming increasingly important [5]. Quality refers to a set of properties and characteristics of a service, which give it the ability to satisfy customer needs. If the company undertakes to deliver the cargo to its destination and on time and in accordance with the terms of safety, in the future, the client expects the carrier to reduce downtime, reduce storage charges, expand the delivery network, etc., i.e. improve the quality of services provided [6] [7].

It is a mistake to state that the provision of quality services is very expensive. On the contrary, non-fulfillment of the terms of the agreement entails additional expenses of material and labor resources aimed at eliminating errors. Thus, systematic disruptions in the transportation schedule ultimately lead to the loss of customers, reputation and a place in the transport services market [5] [6] [8].

Practice shows that in 80% of cases, problems with the quality of transport services are connected with the loss of management by the company. How to ensure quality? The answer to this question may be a model of the quality system included in the overall management system. The ISO 9000: 2000 standard defines a quality system as "the totality of the organizational structure, procedures, processes and resources necessary for the implementation of administrative quality management"

One of the important issues of logistics services is the price, as expected compensation for the total package of services that the firm offers to the consumer. It is much more difficult to determine the price of logistics services than the price of transportation itself, since in many respects the price of logistics services depends on the customer's perception of the entire service system. The choice of the optimal level of customer service is determined by the dynamics of the value of costs. It has been revealed that, starting from 70% and above, maintenance costs grow exponentially depending on the level of service, and at a service level of 90% and above, the service becomes unprofitable [8].

The work of transport should be based on customer requests. The client is attracted by the minimum delivery times, 100% safety of the goods during transportation, convenience of receiving and delivering goods, the possibility of obtaining reliable information about tariffs, conditions of transportation and location of the goods. Only under these conditions, the client is ready to bear the appropriate costs.

The above is of particular theoretical and practical interest in the study of patterns of change in profitability (profitability) of road transport enterprises (Urta Osie Trans JSC) from their performance indicators for 2015-2018.

In this paper, we will consider two ways to apply linear regression in the study of indicators of the organization of freight traffic, namely: steam and multiple regression. That will allow to calculate the potential profit of the enterprise, based on the data already available.

Research design

Linear regression - a regression model used in statistics for the dependence of one (explained, dependent) variable Y on another or several other variables (factors, regressors, independent variables) X with a linear dependence function.

Regression model:

$$y = f(x, b) + \mathcal{E}, E(\mathcal{E}) = 0 \quad (1)$$

where b - model parameters, $\mathcal{E}$ - random error

Multifactorial regression equation:

$$y = f(x, b) = b_0 + b_1x_1 + b_2x_2 + \ldots b_kx_k \quad (2)$$

Where b_j are regression parameters (coefficients), x_j are regressors (model factors), k is the number of model factors.

Regression is of two types: steam (or two-factor) and multiple (or multifactorial). Such regressions differ from each other in the type of equation and the number of independent variables. Paired regression equations relate to first order regression equations, and multiple regression equations - to non-linear regression equations. The parameters of the linear regression equation y = a + bx are found by the least squares method from the system of normal equations:

$$\begin{cases} a*n + b*\sum x = \sum y \\ a*\sum x + b*\sum x^2 = \sum xy \end{cases} \quad (3)$$

Results and discussion

For example, we will find the coefficients a and b, having calculated the paired linear regression for the indicators "Cargo turnover in km." (We take it for X) and "Income from traffic" (we take for Y) in Table 1:

Table 1. Indicators needed to calculate steam regression

Year	∑X	∑Y	$\sum X^2$	∑XY
2015	47300	10320	2237290000	488136000
2016	24523,1	9392,5	601382434	230333217
2017	41317	10831	1707094489	447504427
2018	27485	10684,3	755425225	293657986
∑	140625,1	41227,8	5301192148	1459631629

Based on these data, we compose a system of linear equations:

$$\begin{cases} a*4 + b*140625,1 = 41227,8 \\ a*140625,1 + b*5301192148 = 1459631629 \end{cases} \quad (4)$$

After solving, we get the following results: b = 0.028588535, a = 9301.883596. We substitute the obtained values into the formula y = a + bx, taking the turnover of goods in the amount of 50,000 tkm for X. Profit will be:

$$\begin{aligned} y = \ & b_0 + bx = 9301,883596 \\ & +0,028588535*50000 \\ & = 10731 \text{ mln.soums} \end{aligned} \quad (5).$$

Let us apply this formula in turn to the remaining characteristics in order to track their impact on the final financial result (profit). To do this, alternately, for X, we will take each of the indicators presented by B as the initial data, and Y will give the constant value of the indicator "Income from transportation".

Table 2. The coefficients a and b for all given operational ATP indicators

	X	Y	X2	XY	a	b
Total load	15917	41227,8	64572233	163824517,7	11052,20	-0,19
Park utilization rate	2,66	41227,8	1,7836	27544,51	10165,83	282,24
Total load	2,15	41227,8	1,1617	22242,243	9832,32	949,26
Tonnage utilization rate	2,98	41227,8	2,2444	30550,11	10416,04	-218,19
Working day mode	39	41227,8	381	402885,5	-1580,90	1219,27
Average transportation distance	903,3	41227,8	229530,69	9448061,55	9092,34	5,38
Downtime under loading and unloading	91,4	41227,8	2143,96	949919,5	7537,84	121,72
Average daily mileage	490	41227,8	60700	5077839	5328,28	40,64
Technical speed	200,2	41227,8	10020,04	2063268,5	11217,76	-18,22
Transportation volume in tons	684,7	41227,8	128140,09	6956519,35	11863,61	-9,10
Cargo handling in t.km	140625,1	41227,8	5301192148	1459631629	9301,89	0,03
Parking days in household	168470	41227,8	7241050506	1734009275	11002,35	-0,02
Parking days at work	111613	41227,8	3197535447	1154483813	8933,36	0,05
Downtime	50404	41227,8	735764314	509362351,7	11577,92	-0,10
Hours on duty	1092,5	41227,8	308795,25	11323844,25	8662,61	6,02
Motion hours	298,9	41227,8	24315,21	3105390,65	9410,13	12,04
Downtime hours	792,6	41227,8	161462,76	8207769,3	8624,25	8,50
Total mileage	14972,2	41227,8	60909569,64	155540628,3	9366,78	0,25
Mileage with load	8167,9	41227,8	18550609,81	85068537,05	9345,46	0,47
Rate of readiness	2,81	41227,8	2,0137	29172,853	10134,19	345,53
Production on the machine day	1513,4	41227,8	587197,38	15601090,57	10241,51	0,17

** In this table, red indicates various coefficients and constant values, indicators of which can distort the overall picture. And green - indicators, the increase of which should have a positive impact on the total profit of the enterprise.*

For ease of computation and working with large amounts of data, we can use the library for the Python programming language. In this library, there is a functional necessary to solve systems of linear equations.

To do this, it is only necessary to transfer the functions of two matrices: X, all the technical indicators necessary in comparison, and Y, transportation revenues for 2015, 2016, 2017, 2018. At the output, we obtain the coefficients a and b for each indicator Table 2.

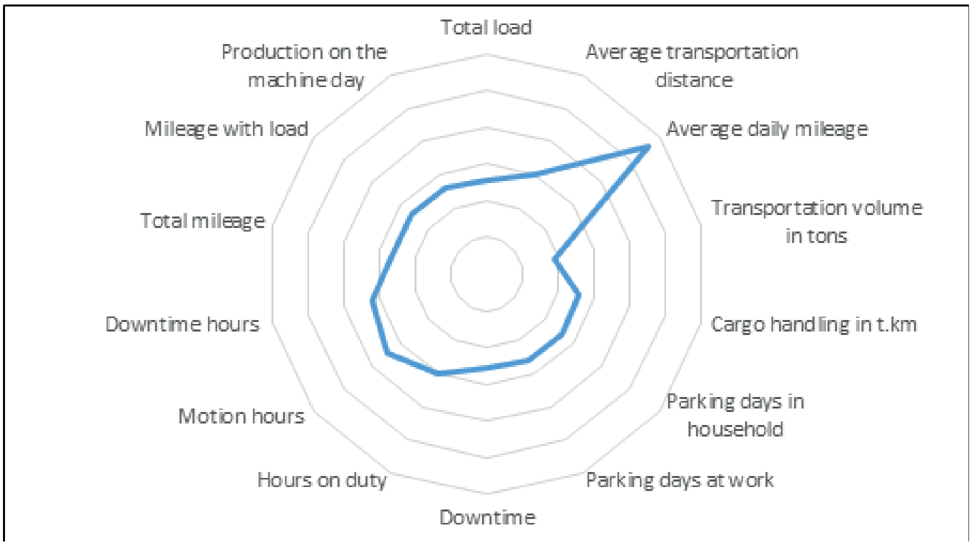
However, with a large amount of input data, this method is not optimal, and shows only special cases of correlation of indicators. Therefore, let us return to formula (2) and apply it to multifactorial regression:

$$y = f(x,b) = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 + b_8x_8 + b_9x_9 + b_{10}x_{10}$$
$$+ b_{11}x_{11} + b_{12}x_{12} + b_{13}x_{13} + b_{14}x_{14} + b_{15}x_{15} + b_{16}x_{16} + b_{17}x_{17} + b_{18}x_{18} + b_{19}x_{19} + b_{20}x_{20} + b_{21}x_{21} \quad (6)$$

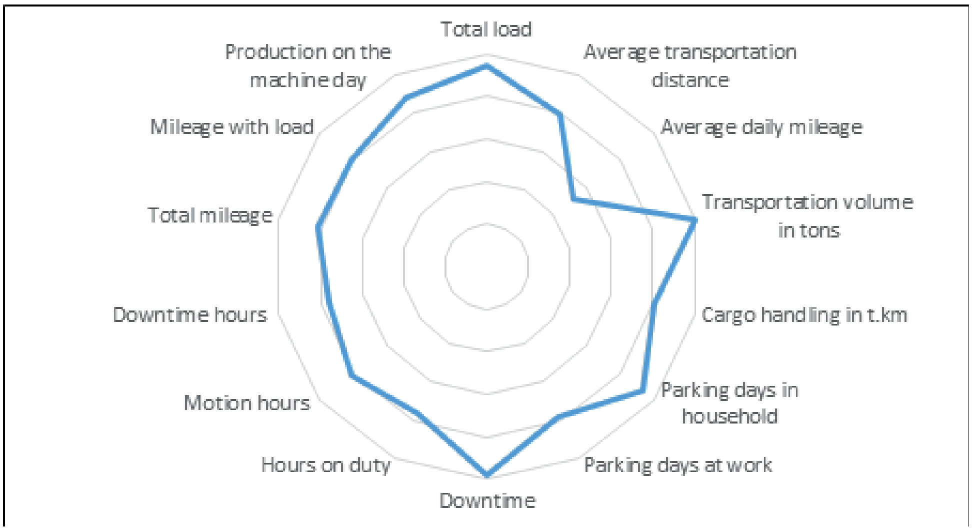
Table 3. Systems of linear equations built to obtain values from Table 2.

Total load	a*4+b*15917.0=41227.8
	a*15917.0+b*64572233.0=163824517.7
Park utilization rate	a*4+b*2.66=41227.8
	a*2.66+b*1.7836=27544.51
Mileage utilization rate	a*4+b*2.15=41227.8
	a*2.15+b*1.1617=22242.243
Tonnage utilization rate	a*4+b*2.98=41227.8
	a*2.98+b*2.2444=30550.11
Working day mode	a*4+b*39.0=41227.8
	a*39.0+b*381.0=402885.5
Average transportation distance	a*4+b*903.3=41227.8
	a*903.3+b*229530.69=9448061.55
Downtime under loading and unloading	a*4+b*91.4=41227.8
	a*91.4+b*2143.96=949919.5
Average daily mileage	a*4+b*490.0=41227.8
	a*490.0+b*60700.0=5077839.0
Technical speed	a*4+b*200.2=41227.8
	a*200.2+b*10020.04=2063268.5
Transportation volume in tons	a*4+b*684.7=41227.8
	a*684.7+b*128140.09=6956519.35
Cargo handling in tkm	a*4+b*140625.1=41227.8
	a*140625.1+b*5301192148.0=1459631629.0
Parking days in household	a*4+b*168470.0=41227.8
	a*168470.0+b*7241050506.0=1734009275.0
Parking days at work	a*4+b*111613.0=41227.8
	a*111613.0+b*3197535447.0=1154483813.0
Downtime	a*4+b*50404.0=41227.8
	a*50404.0+b*735764314.0=509362351.7
Hours on duty	a*4+b*1092.5=41227.8
	a*1092.5+b*308795.25=11323844.25
Motion hours	a*4+b*298.9=41227.8
	a*298.9+b*24315.21=3105390.65
Downtime hours	a*4+b*792.6=41227.8
	a*792.6+b*161462.76=8207769.3
Total mileage	a*4+b*14972.2=41227.8
	a*14972.2+b*60909569.64=155540628.3
Mileage with load	a*4+b*8167.9=41227.8
	a*8167.9+b*18550609.81=85068537.05
Rate of readiness	a*4+b*2.81=41227.8
	a*2.81+b*2.0137=29172.853
Production on the machine day	a*4+b*1513.4=41227.8
	a*1513.4+b*587197.38=15601090.57

Solve this equation in the same way using the Python programming language. The data obtained will be listed in Table 3.



a)



b)

Figure 1. Visualization of coefficients of a and b from Table 3.

On these diagrams, the following dependence can be traced: the more the indicator is knocked out of the general range, the greater the positive impact its growth will have on the total profit.

Let us check the data, substitute for them the indicators for 2015 that are already known to us, to find out how accurate the forecast is:

$$\begin{aligned} y &= 10392 + 4104*0.00274 + 0.7*0 + 0.56*0 + 0.7*0 + 10*0.00001 + 315*0.0003 + 25*0.00012 + \\ &+ 130*0.0004 + 50*0 + 150*-0.00123 + 47300*-0.04776 + 43947*0.01571 + 30763*0.09617 + \\ &+ 10415*-0.13752 + 308*0.00129 + 104*-0.00014 + 204*0.00142 + 5200*-0.00736 + 2912*- \\ &- 0.00122 + 0.76*0 + 335.5*0.00021 = 10319 \end{aligned} \quad (7)$$

Since the revenue for 2015 amounted to 10320, which ultimately gives an error of 0.0029%.

Table 4. The share of the impact of the operational indicator on the final financial result according to the multifactorial regression equation.

	Indicators	Shares of parameters' impact in financial results
1.	Working day mode	0,008448%
2.	Downtime under loading and unloading	0,101377%
3.	Production on the machine day	0,17741%
4.	Average transportation distance	0,253443%
5.	Average daily mileage	0,337923%
6.	Motion hours	1,089803%
7.	Downtime hours	1,199628%
8.	Total load	2,314776%
9.	Parking days in household	13,27194%
10.	Parking days at work	81,24525%

Figure 2 shows that in a comprehensive review of all the coefficients involved in the formation of the total profit in Table 5, the most important role is played by the parameters "Machine days at work", "Machine days in household".

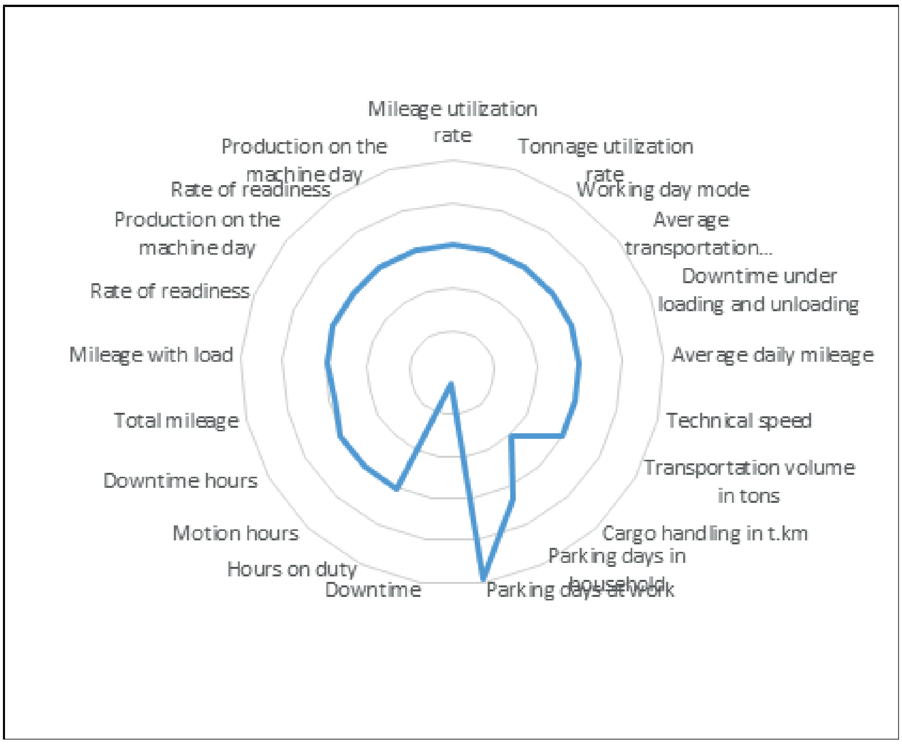


Figure 2. Visualization of coefficients under multifactorial regression.

Table 5. The necessary coefficients to solve the multifactorial regression equation.

b1	10392
b2	0,00000
b3	0,00000
b4	0,00000
b5	0,00001
b6	0,00030
b7	0,00012
b8	0,00040
b9	0,00000
b10	-0,00123
b11	-0,04776
b12	0,01571
b13	0,09617
b14	-0,13752
b15	0,00129
b17	-0,00014
b18	0,00142
b19	-0,00736
b20	-0,00122
b21	0,00000

Conclusion

In conclusion, it can be noted that, based on the analysis and practical application, there are two economic and mathematical models: steam and multivariate regression, to determine the potential profit of the motor company, depending on the performance of its work for 4 years including 2015, 2016, 2017, 2018. It was found that the use of pair regression is the most important indicator in the ATP is the "Average daily mileage". And when considering all the indicators in the amount, with multifactorial regression, the greatest influence is exerted by the indicators "Machine - days at work" and "Machine - days at the households".

It is with the above indicators that there is a maximum correlation of the profit of the automobile company. Also, except for "Machine - days at work" and "Machine - days at the farm" other indicators have a positive effect on profit in Table 4.

Thus, using this approach, it is possible to determine the potential profit of an enterprise on the basis of known input data or their predicted change.

References

1. A. Judith, B. Thomas i E. Hans, „Managing Complexity in Supply Chains: A Discussion of Current Approaches on the Example of the Semiconductor Industry," Procedia CIRP, pp. 79-84, 2013.

2. C. Cheng-Yang, C. Tzu-Li i C. Yin-Yann, „An analysis of the structural complexity of supply chain networks," Applied Mathematical Modelling, tom 38, nr 9-10, pp. 2328-2344, 2014.

3. S. Can i R. Thomas, „Supply Chain Complexity in the Semiconductor Industry: Assessment from System View and the Impact of Changes," IFAC-PapersOnLine, tom 48, nr 3, pp. 1210-1215, 2015.

4. M. Ignas i C. Jonas, „Dynamic Capabilities in Supply Chain Management," Procedia - Social and Behavioral Sciences, tom 213, pp. 830-835, 2015.

5. D. Luigi, I. Angel i C. Patricia, „The quality of service desired by public transport users," Transport Policy, tom 18, nr 1, pp. 217-227, 2011.

6. K. Milorad, „Measurement of logistics service quality in freight forwarding companies: A case study of the Serbian market," The International Journal of Logistic Management, tom 27, nr 3, 2014.

7. S. Oksana i K. Irina , „Transport Infrastructure Development Performance," Procedia Engineering, p. 319 – 329, 2017.

8. L. B. Mirotin, „Transport logistics: A textbook for transport universities," 2003.