

Wojciech Kamiński

Silesian University of Technology

e-mail: wojciech.kaminski@polsl.pl

**DETERMINATION THE IMPACT OF SOCIO-ECONOMIC FACTORS
ON THE VOLUME OF PASSENGER TRANSPORT ON SELECTED
RAILWAY LINES IN POLAND**

ABSTRACT

Background: The volume of transport on a given railway line depends on many socio-economic factors. Therefore, before starting the modernization of the line, various analyzes are carried out regarding the scope of work. During such analyzes, many socio-economic factors having a different impact on the volume of passenger and freight rail transport should be taken into account. Many factors have been taken into account so far, but their weights were determined by decision centers or experts.

Methods: The weights of individual socio-economic factors were determined using correlation coefficients. Individual weights were determined, taking into account the impact of various factors on the volume of transport. During the calculations, an analysis was made on railway lines located in various regions of Poland. Two lines were selected from each voivodeship, according to the assumption that one of them is the main line and the other is a local line.

Results: Weights were obtained determining the impact of individual analyzed factors on the volume of transport. This analysis allowed to eliminate factors that had no significant impact on rail transport.

Conclusions: The weights determined in this way using the previously calculated correlation coefficients will be able to be used in the future when analyzing railway lines in Poland in terms of their use in passenger transport, thus it will be possible to avoid subjective expert assessments.

Keywords: railway transport, passenger transport, correlation coefficient, logistics

INTRODUCTION

Currently, many railway lines in Europe and Poland are being modernized or revitalized. In the case of construction of a new line, an analysis should be made to determine how its construction will change the transport accessibility of the area where it will be located [Sanchez-Mateos and Givoni 2012]. The rail passenger transport system is very complex [Abril et al. 2008]. Therefore, during making decisions on its functioning, numerous multi-criteria analyzes are performed, allowing for consideration of various economic, environmental and social factors [Perez et al. 2015]. The current analyzes of this type have taken into account various factors determining their weights based on expert assessments or through discussions with decision-making centers [Jacyna and Wasiak 2008].

The purpose of this article is to determine the weights of individual socioeconomic factors affecting the volume of passenger transport on individual railway lines using correlation coefficients. Taking into account the impact of a particular factor on the volume of transport, it is possible to obtain a weight to it ranging from 0 to 1. The weight values determined in this way can be used in the future for analyzing railway lines in Poland in terms of their use in passenger transport.

Calculations were carried out for railway lines located in various regions of Poland. Two lines were selected from individual voivodships: one main line and the other local line. In total, 32 railway lines were used for the analysis. Correlation coefficients were calculated for various factors affecting the volume of transport. Gauges describing the volume of regional and long-distance transport were used. Analysis of correlation coefficients determining the impact of selected socio-economic factors on the size of gauges describing rail passenger transport in Poland has already been performed. However, during previous analyzes, the focus was on comparing the volume of passenger rail transport in individual voivodships, choosing other factors for this purpose and without analysis individual railway lines [Kamiński and Sładkowski 2019a]. Other analysis was made in the region of Łódź, where calculations were limited to 4 railway lines and only regional traffic was included [Kamiński and Sładkowski 2019b]. Taking into account the larger number of railway lines located in different parts of Poland allowed to increase the reliability of the results by increasing the sample, which is important during statistical calculations. The obtained weights reflect the transport throughout the country, not limited to the local specificity of one region.

DATA USED FOR ANALYSIS

To calculate the correlation coefficients, gauges describing the volume of rail passenger transport in individual voivodships in Poland were used. The gauges taken into account are: transport work done in regional traffic and transport work done in long-distance traffic. The analysis used gauges and socio-economic factors affecting the volume of transport that occurred in Poland in 2017. During calculation of correlation coefficients, these gauges were used as explanatory (dependent) variables. Long-distance traffic is carried out by the Polish National Railways Intercity carrier on all analyzed railway lines. However, there are various regional carriers on individual lines. Most lines operate Polregio. Other regional carriers are: Lower Silesia Railways, Arriva (Kuyavian-Pomeranian Voivodeship), Łódź Agglomeration Railways, Mazovian Railways, Silesian Railways and Greater Poland Railways. All analyzed gauges occurring on the analyzed railway lines are presented in Table 1. Main railway lines are bold. Due to the fact that the analyzed railway lines are of different lengths, to make possible calculations, the values of individual gauges were not used, but their relation to the length of the lines. The length of these lines is also shown in the table. In this way, instead of the value of transport work done, the number of launched trains on a line per year (in regional and long-distance traffic) was obtained.

The volume of passenger transport on a particular railway line depends on the distribution of population centers and the transport needs of the area through which the railway line passes. Socio-economic factors affecting the volume of rail passenger transport on individual lines were taken into account, such as: population of cities along the railway line, number of registered vehicles per 1000 inhabitants, number of business entities having split into the size of companies (expressed in the number of employees), availability of inhabitants to the railway line, number of beds in accommodation facilities in communes at the railway line, number of commuters to work and average number of bus and tram connections from city center to the railway station. These data are presented in Table 2. The source of data are reports published by the Central Statistical Office. One of them, availability of inhabitants to the railway line, have been own calculated. To enable comparison of analyzed railway lines, which have different lengths during the calculation of correlation coefficients, the values of individual factors were not used directly, but their ratio to the length of the line.

CHALLENGES AND MODERN SOLUTION IN TRANSPORTATION

Table 1. Gauges describing the volume of transport

Section:	Transport work done in regional traffic [train-km]:	Transport work done in long-distance traffic [train-km]:	Line length [km]:
Wrocław Gł. – Legnica	1108250	194870	65
Kłodzko Gł. – Wałbrzych Gł.	239802	0	51
Bydgoszcz Gł. – Toruń Gł.	741540	728790	51
Grudziądz – Chełmża	298224	0	38
Lublin – Rejowiec	618200	260590	55
Zamość Wsch.– Zwierzyniec	199360	79936	32
Zbąszynek – Rzepin	352050	921150	75
Żary – Zielona Góra	233094	0	53
Łódź Kaliska - Sieradz	731534	401436	59
Łódź Kaliska - Kutno	481206	248200	68
Kraków Gł. – Tarnów	1408946	8819118	77
Nowy Sącz – Muszyna	386150	62016	51
Warszawa Zach. – Grodzisk Maz.	1565136	992736	27
Nasielsk – Sierpc	321200	0	88
Opole Gł. – Brzeg	504800	567200	40
Nysa – Kędzierzyn Koźle	452100	0	75
Rzeszów Gł. – Przeworsk	528048	415368	36
Jasło – Zagórz	292008	180780	69
Białystok – Sokółka	326508	71316	42
Białystok – Czeremcha	224840	0	77
Gdańsk Gł. – Tczew	477056	487232	32
Chojnice – Kościerzyna	237084	0	69
Katowice – Gliwice	847908	515376	27
Żywiec – Zawadoń	331298	5698	37
Kielce – Skarżysko-Kamienna	379080	493110	45
Skarżysko-Kamienna – Ostrowiec Świętokrzyski	267812	87124	46
Olsztyn Gł. – Iława Gł.	589260	705180	69
Olsztyn Gł. – Braniewo	290496	0	96
Poznań Gł. – Zbąszyń	908240	819624	74
Leszno – Wolsztyn	211140	0	46
Szczecin Gł. – Goleniów	699276	334932	38
Kołobrzeg – Goleniów	542400	45200	100

Source: own data

CALCULATION OF AVAILABILITY OF INHABITANTS TO THE RAILWAY LINE

One of the factors affecting the volume of passenger transport on a railway line was not obtained directly from the reports of the Central Statistical Office. The parameter of availability to the railway line for individual residents is a value related to the location of stations (or stops) in population centers (cities and villages). Its size is influenced by the distance of the station from the center and the size of the city in which it is located.

CHALLENGES AND MODERN SOLUTION IN TRANSPORTATION

Table 2. Factors affecting the volume of transport

Section:	Popul. [pe.]:	No. of veh. / 1000 pe.:	Bus. ent. up to 9 pe.:	Bus. ent. 10 - 49 pe:	Bus. ent. over 50 pe.:	Av. of inh. to the line. [pe./km]:	No. of beds [pe.]:	No. of comm. [pe.]:	Av. no. of con. in a day [con.]:
Wroc. Gł. – Legnica	758013	637.22	135152	3641	985	3591.72	13296	73076	285.62
Kłodz. Gł. – Wałbrz. Gł.	184290	557.68	20654	577	141	1185.58	2601	12719	8.33
Bydg. Gł. – Toruń Gł.	572163	559.15	68354	2330	698	3516.38	8115	43327	210.00
Grudziądz – Chełmża	111431	523.75	9684	369	115	1954.70	1425	3813	31.33
Lublin – Rejowiec	388645	579.17	48982	1514	418	2735.39	4526	41053	34.86
Zam. Wsch. – Zwierzyn.	76677	519.60	8541	279	90	1919.04	2544	7084	23.89
Zbąszynek – Rzepin	40853	634.42	6528	227	54	356.75	2566	6295	2.64
Żary – Ziel. Góra	185775	627.69	26129	897	208	1514.98	1534	21145	28.10
Łódź Kal. – Sieradz	863521	728.76	107563	4303	925	10267.10	9521	57623	74.15
Łódź Kal. – Kutno	828287	718.48	102150	3985	903	10864.94	8965	53155	86.56
Kraków Gł. – Tarnów	948204	587.11	163658	6068	1421	5268.79	35440	123357	80.00
Nowy Sącz – Muszyna	104896	447.03	13308	527	143	987.71	4755	11364	69.13
War. Zach. – Gro. Maz.	2025760	704.49	442698	14809	3939	20361.57	34149	283173	402.50
Nasielsk – Sierpc	64274	720.01	7012	270	65	456.27	611	3274	2.86
Opole Gł. – Brzeg	174910	663.34	27909	882	237	2478.87	1839	24582	71.63
Nysa – Kędz. Koźle	132923	595.81	19273	654	159	1462.31	5057	9562	25.46
Rzesz. Gł. – Przeworsk	241271	521.78	31858	1018	318	3633.26	4469	52132	45.90
Jasło – Zagórz	145984	526.26	17548	667	221	1821.32	2831	32628	26.91
Białystok – Sokółka	337044	468.84	38361	1134	386	3542.62	3094	17521	59.78
Białystok – Czeremcha	328781	479.99	38718	1155	399	1837.33	2863	17894	46.86
Gdańsk Gł. – Tczew	564995	592.58	85415	2824	670	8287.99	18557	49672	178.00
Chojnice – Kościerzyna	80769	589.98	7976	305	92	609.45	905	5758	21.79
Katowice – Gliwice	948112	568.43	107774	4971	1281	20223.53	8960	174321	327.50
Żywiec – Zwardoń	60542	579.21	7308	306	74	1032.82	1682	9718	32.33
Kielce – Skarż.-K.	254824	526.39	36147	1317	372	2843.10	5726	29370	40.27
Skarż.-K. – Ostr. Św.	176108	503.15	18966	639	195	2677.22	1379	13415	11.00
Olsztyn Gł. – Iława Gł.	244163	533.92	29237	997	291	1914.11	8099	21784	84.23
Olsztyn Gł. – Braniewo	214861	531.88	28478	983	249	1258.51	6091	19930	62.33
Poznań Gł. – Zbąszyń	584872	660.13	121611	4568	977	2842.12	10617	98209	197.87
Leszno – Wolsztyn	88347	605.79	15812	725	167	1125.74	3706	13128	9.15
Szczec. Gł. – Goleniów	428136	545.06	71782	2207	459	5199.20	7624	26800	110.00
Kołobrzeg – Goleniów	120852	583.26	20622	627	160	872.65	23929	8174	19.94

Source: reports of the Central Statistical Office and own calculations

If the station is located in the city center, the parameter related to its location takes the value 1, while if the travel time from the city center to the station is 10 minutes or more (taking into account distance, the condition of roads and occurring traffic congestion) it sets the value to 0. In other cases, proportional value is assumed based on the time of arrival to the stop. The size of the city in which the station is located is also taken into account. The availability of a particular station is a product of its location and the size of the city where it is located. In order to determine accessibility to the entire railway line, the availability of all stops located on it is then added. Later, the quotient of the sum of the availability of all stops and the length of the analyzed railway line is calculated [Kamiński and Sładkowski 2019b]. Locating new stops on the line can only increase its availability to residents, cannot lower it. The calculation of the availability of inhabitants to the railway line is shown by formula (1).

$$A_L = \frac{\sum (S_L \cdot C_S)}{L_L} \quad (1)$$

where:

A_L - availability of inhabitants to the railway line

S_L - location of the station or stop

C_S - size of the city/village where the station/stop is located

L_L - length of the analyzed railway line.

CALCULATION OF CORRELATION COEFFICIENTS

The term of correlation determines the interrelationships between specific selected variables. During characterizing the correlation between two variables its direction and strength should be given. During the analysis of the impact of socio-economic factors on passenger services performed on individual railway lines, the strength of the correlation and its level of significance are more important than the direction. If the absolute value of the correlation coefficient is less than 0.2, then there can be no relationship between the variables, because the correlation is very weak. If a higher value of the correlation coefficient is obtained, a smaller sample size is enough to ensure its significance [Altman and Krzywinski 2015]. There are several types of correlation coefficients.

Pearson correlation coefficient is used to test linear relations between two variables, when increasing the value of one variable is accompanied by proportional changes in the value of the second variable [Sedgwick 2012]. It is the simplest, however, during calculating the Pearson correlation coefficient between two variables, the calculated value can be significantly

disturbed due to possible relationships of other variables with the analyzed pair of variables [Wang et al. 2018].

The partial correlation between two variables is the interrelationship between these variables after excluding the influence of other variables on the analyzed ones. To calculate this coefficient is used a correlation matrix, containing interrelations between all analyzed variables [Qian et al. 2015]. Calculations of partial correlation coefficients are usually complicated and calculation software is used to perform them.

Calculations of weights of individual factors affecting the volume of passenger rail transport on selected railway lines located in Poland were made using Pearson correlation coefficients and partial correlation. Statistical calculations software STATISTICA was used to do them. The calculated values of Pearson and partial correlation coefficients for socio-economic factors affecting the volume of transport on individual railway lines in Poland are presented in Table 5. This table also indicates the significance of the obtained correlation coefficients. Bold correlation coefficients indicate significant correlations at 0.01 bilaterally, while italics indicate significant correlations at 0.05 bilaterally. Other correlation coefficients are not statistically significant.

Table 5. Calculated values of correlation coefficients

Factor:	Transport work done in regional traffic/line length [trains]:		Transport work done in long-distance traffic/line length [trains]:	
	Pearson correlation:	Partial correlation:	Pearson correlation:	Partial correlation:
Population [pe./km]	0.960	0.257	<i>0.363</i>	0.098
Number of vehicles /1000 pe.	0.292	-0.082	0.128	0.041
Business entities up to 9 people / km line	0.937	<i>0.308</i>	<i>0.354</i>	-0.222
Business entities 10 - 49 people / km line	0.948	-0.015	<i>0.367</i>	0.240
Business entities over 50 people / km line	0.944	<i>-0.389</i>	<i>0.353</i>	-0.138
Availability of inh. to the line. [pe./km]	0.840	-0.171	0.325	-0.186
No. of beds in accom. fac. [pe./km]	0.896	0.046	<i>0.515</i>	<i>0.509</i>
No. of commuters to work [pe./km]	0.952	<i>0.504</i>	<i>0.365</i>	0.033
Av. no. of con. in a day [con.]	0.855	0.159	0.284	-0.124

Source: own calculations

Due to the very small obtained values of both Pearson and partial correlation coefficients, all statistically insignificant, the factor which is the number of registered vehicles per 1000 inhabitants was omitted during further analysis. This factor does not affect the volume of transport on individual railway lines. This can be seen in Figure 1, which presents a spread chart of the quotient of transport work done in passenger traffic to the length of the line and the

number of vehicles per 1000 inhabitants. This figure shows not only the lack of a linear relationship, but any relationship between these quantities at all.

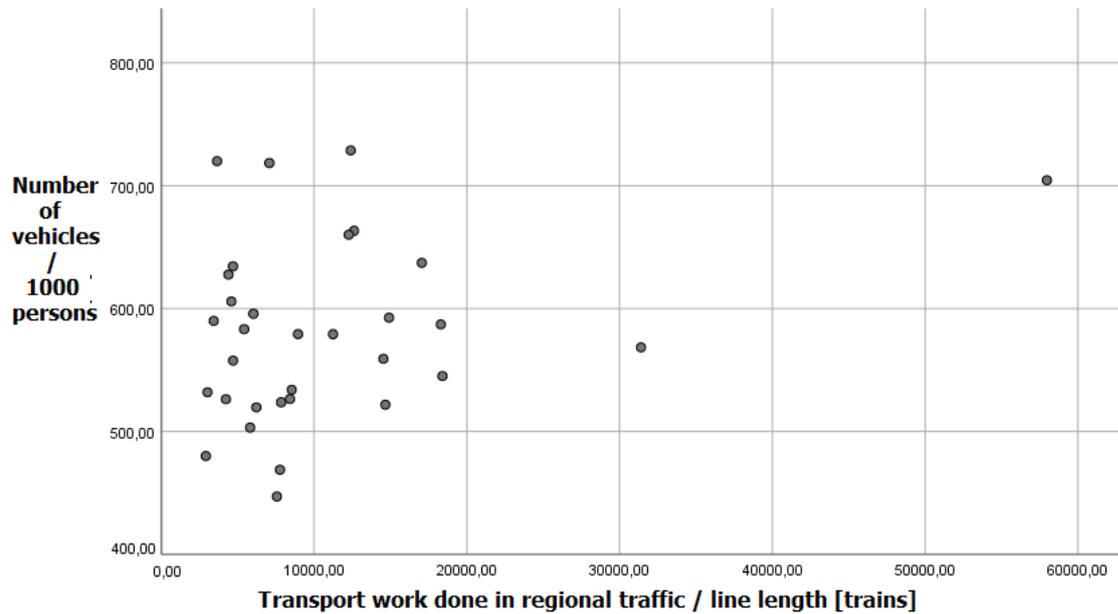


Fig. 1. Spread chart of the transport work done in regional traffic and number of vehicles.

Source: own calculations

CALCULATION OF WEIGHTS OF INDIVIDUAL SOCIO-ECONOMIC FACTORS

During calculating the weights of individual factors affecting the volume of rail passenger transport on selected lines in Poland, the absolute value of correlation coefficients calculated for two gauges describing the volume of transport was added for each socio-economic factor. The use of an absolute value is due to the fact that when determining the weights of various factors, only the strength of their impact on transport is important, not the direction. Later the added sum was divided by the two (number of analyzed gauges). Then, by dividing each of the received values defining the impact of a one factor on the volume of transport by the sum of the impacts of all factors, it is possible to determine the weight. The sum of all obtained factors in this way is 1. The actions performed during calculating the weights are shown in formula (2).

$$W_F = \frac{\sum C_C}{2 \cdot \sum I_F} \tag{2}$$

where:

W_F - obtained weight of individual factor

C_C - calculated correlation coefficients between individual factors and gauges describing the size of transport

I_F – impact of a particular factor on the volume of transport.

The calculation of the sum of the absolute values of the obtained correlation coefficients (I_F) is presented in Table 4.

Table 4. Sum of absolute values of obtained coefficients

Factor:	Sum of correlation coefficients:	
	Pearson correlation:	Partial correlation:
Population [pe./km]	0.662	0.178
Business entities up to 9 people / km line	0.646	0.265
Business entities 10 - 49 people / km line	0.658	0.128
Business entities over 50 people / km line	0.649	0.264
Availability of inh. to the line. [pe./km]	0.583	0.179
No. of beds in accom. fac. [pe./km]	0.706	0.278
No. of commuters to work [pe./km]	0.659	0.269
Av. no. of con. in a day [con.]	0.570	0.142
Sum	5.129	1.700

Source: own calculations

While the calculation of the weights of individual factors by dividing the sum of the received correlation coefficients for two gauges by the sum of the influence of all factors is presented in Table 5. In addition, this table presents the percentage difference in weights obtained using Pearson correlation coefficients and partial correlation coefficients.

Table 5. Calculated weight values and difference between the two methods

Factor:	Calculated weights:		Difference [in%]:
	Pearson correlation:	Partial correlation:	
Population [pe./km]	0,129	0,104	19.38
Business entities up to 9 people / km line	0,126	0,156	19.23
Business entities 10 - 49 people / km line	0,128	0,075	41.41
Business entities over 50 people / km line	0,126	0,155	18.71
Availability of inh. to the line. [pe./km]	0,114	0,105	7.89
No. of beds in accom. fac. [pe./km]	0,138	0,163	15.34
No. of commuters to work [pe./km]	0,128	0,158	18.99
Av. no. of con. in a day [con.]	0,111	0,083	25.23
Sum	1,000	1,000	-

Source: own calculations

The weight values obtained using Pearson's correlation coefficients for all factors are very similar to each other. Larger differences occur in the case of weights obtained using more accurate partial correlation coefficients. Based on the weights obtained using the partial correlation coefficient, it can be stated that the number of beds and the number of small business entities employing up to 9 people have the greatest impact on the volume of transport. Spread

chart with a visible linear relationship for the quotient of transport work done in passenger traffic to the number of small business entities is shown in Figure 2.

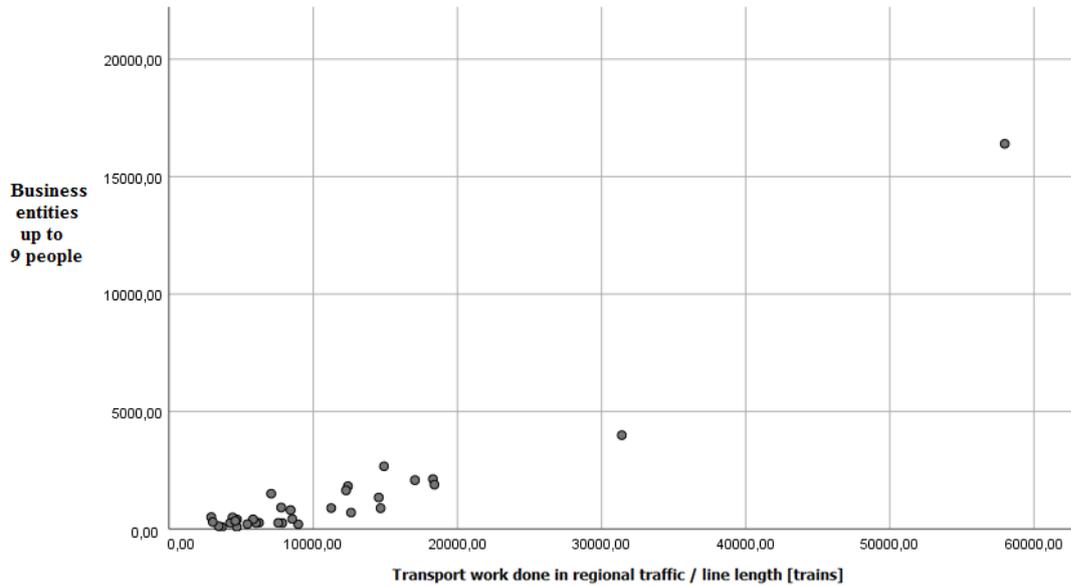


Fig. 2. Spread chart of the transport work done in regional traffic and number of small business entities.

Source: own calculations

In turn, the number of medium-sized enterprises (employing between 10 and 49 employees) and the average number of bus and tram connections, enabling access to the railway station from the city center, have the least impact on the volume of passenger transport. Figure 3 shows spread chart for the quotient of transport work done in passenger traffic to the number of average number of bus and tram connections with a slight dependency of the analyzed variables.

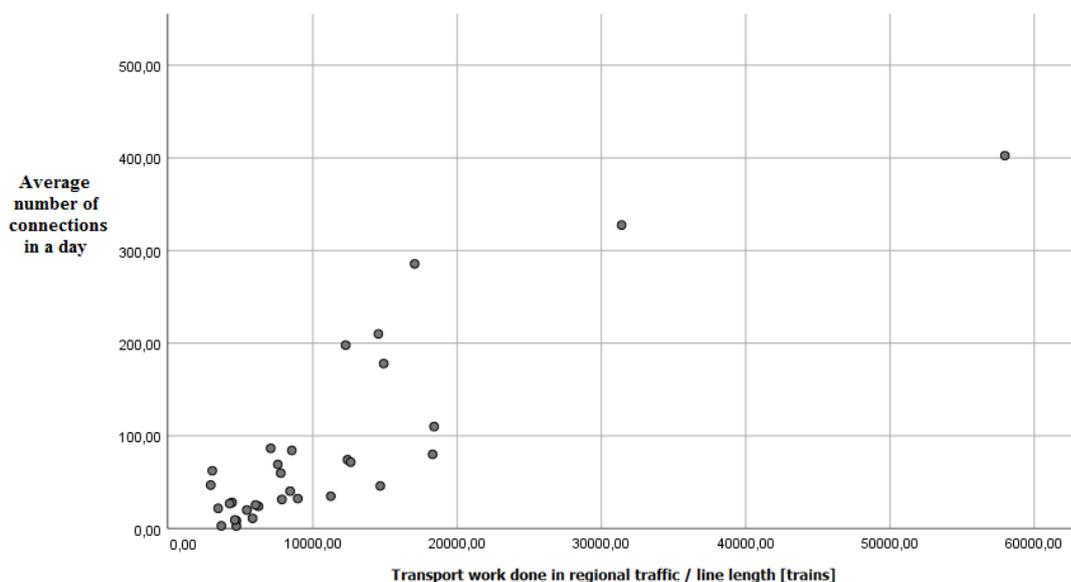


Fig. 3. Spread chart of the transport work done in regional traffic and number of bus and tram connections.

Source: own calculations.

Determination weights of socio-economic factors was not done very often. Comparison was made for the weight values obtained using partial correlation coefficient with the weights used in analysis of railway lines in Lower Silesia [Grulkowski and Zariczny 2010]. This analysis did not explain how the weights determining the influence of particular factors were obtained. They were probably identified by the authors based on their experience or using one of the expert methods. During this analysis, the weight of 0.25 (0.104 in correlation analysis) for the population was determined. In turn, the number of business entitles assigned a weight of 0.12 (in this analysis after adding up the impact of all business units 0.386). The next parameter, the accessibility to rail services, was defined slightly differently, as the share of the population living up to 1km from the station in relation to the entire population. The value of this parameter was set to 0.20 (0.105 in this analysis). Other factors taken into account in both analyzes were different. However, in the case of similar factors, it can be stated that the weights obtained using the partial correlation coefficient are completely different from the weights determined during the analysis of railway lines in Lower Silesia.

CONCLUSIONS

The obtained results of the correlation coefficients between the factors affecting the volume of transport and the gauges describing the volume of passenger transport on individual railway lines allowed to eliminate one of the analyzed factors, the number of vehicles per 1000 inhabitants, which had no impact on rail transport. Merely the use of partial correlation coefficient allowed to determine the weights of socio-economic factors affecting the volume of passenger transport on selected railways located in different regions of Poland, because using Pearson's correlation coefficients, the weights of all factors are similar. Use of the Pearson correlation coefficient due to the association of other variables with the analyzed pairs of variables, the obtained value are significantly disturbed. This disturbance causes that the obtained weight values using both correlation coefficients are very different from each other. Only the weights obtained using the partial correlation coefficient are reliable and can be used in further analysis of railway lines in Poland.

REFERENCES

- Abril M., Barber F., Ingolotti L., Salido M.A., Tormos P., Lova A., 2008, An assessment of railway capacity, *Transportation Research Part E*, 44, 774–806, DOI: 10.1016/j.tre.2007.04.001.

- Altman N., Krzywinski M., 2015, Association, correlation and causation, *Nature Methods*, 12, 10, 899–900, DOI: 10.1038/nmeth.3587.
- Grulkowski S., Zariczny J., 2010, Analiza celowości i możliwości rewitalizacji linii kolejowych w województwie Dolnośląskim [Analysis of the purpose and possibilities of revitalization of railway lines in the Lower Silesian Voivodship], *Transcomp – XIV International Conference, Zakopane*, 1025-1038.
- Jacyna M., Wasiak M., 2008, Zastosowanie wielokryterialnej oceny do wyboru wariantu modernizacji elementów infrastruktury kolejowej [Application of multicriteria assessment to the choice of a variant of modernization of railway infrastructure elements], *Problemy Kolejnictwa*, 146, 27–35.
- Kamiński W., Sładkowski A., 2019a, Determination the influence of socio-economic factors on the volume of railway passenger transport in different regions of Poland, *Transport Problems 2019, XI International scientific conference, Katowice-Bochnia, Silesian University of Technology*, CD, 308-319.
- Kamiński W., Sładkowski A., 2019b, Analysis of the influence of socio-economic factors on the volume of railway passenger transport in Łódź region, *The 19th International Multi-Conference RelStat' 19 Abstracts, Riga, Transport and Telecommunication Institute*, 41-42.
- Perez J.C., Carrillo M.H., Montoya-Torres J.R., 2015, Multi-criteria approaches for urban passenger transport systems: a literature review, *Analysis of Operations Research*, 226, 1, 69–87, DOI: 10.1007/s10479-014-1681-8.
- Qian, X. Y., Liu, Y. M., Jiang, Z. Q., Podobnik, B., Zhou, W. X., & Stanley, H. E., 2015, Detrended partial cross-correlation analysis of two nonstationary time series influenced by common external forces, *Physical Review E*, 91, 6, DOI: 10.1103/PhysRevE.91.062816.
- Sanchez-Mateos H.S.M., Givoni M., 2012, The accessibility impact of a new High-Speed Rail line in the UK-a preliminary analysis of winners and losers, *Journal of Transport Geography*, 25, 105–114, DOI: 10.1016/j.jtrangeo.2011.09.004.
- Sedgwick P. 2012, Pearson's correlation coefficient, *British Medical Journal*, 345, DOI: 10.1136/bmj.e4483.
- Wang, G.J., Xie, C. & Stanley, H.E., 2018, Correlation Structure and Evolution of World Stock Markets: Evidence from Pearson and Partial Correlation-Based Networks, 51, 3, 607–635, DOI: 10.1007/s10614-016-9627-7.



Papers included in this monograph was presented during WSLFORUM 2019 conference 18th - 19th of November 2019 in Poznan School of Logistics.



ISBN 978-83-66017-85-6

