Svetla STOILOVA*, Radina NIKOLOVA
Technical University of Sofia
Boulevard Kliment Ohridski 8, Sofia, Bulgaria
**Corresponding author*. E-mail: stoilova@tu-sofia.bg

CLASSIFYING RAILWAY PASSENGER STATIONS FOR USE TRANSPORT PLANNING – APPLICATION TO BULGARIAN RAILWAY NETWORK

Summary. A methodology for the classification of railway passenger stations was developed in this study. Four groups of factors are defined to study the characteristics of the station: potential of the town, importance of the town, infrastructural factors, and characteristics of passengers. In the research we investigated 18 factors and studied 98 passenger stations of railway network in Bulgaria. The method of principal components has been applied for grouping the factors and cluster analysis has been applied to classify the stations. The factors have been classified into 4 components by the method of principal components. The stations have been classified into 6 groups using hierarchical cluster analysis. The methods, Average linkage between group and Within-groups linkage, and distance-type measures Euclidean distance and Squared Euclidean distance were compared to verify the results of cluster analysis. The grouping of the stations has been used for the determination of the stops for categories intercity passenger trains. The main groups of stations for servicing express trains are the first, second, third, fourth and fifth groups. In the sixth group are stations for servicing fast passenger trains. The methodology can be applied to the study of all stations and stops in the rail network.

CLASSIFICATION DES GARES DE CHEMIN DE FER DE PASSAGERS CONCERNANT L'EMPLOI DE PLANIFICATION DES TRANSPORTS – APPLICATION POUR LE RÉSEAU DE FER DE LA BULGARIE

Sommaire. Dans cette étude est élaborée la méthodologie de classification des gares de passagers. On a déterminé 4 groupes de facteurs pour l'étude des caractéristiques des gares: le potentiel de la ville, importance de la ville, facteurs infrastructurels et caractéristiques des passagers.18 facteurs sont étudiés au total. On a étudié 98 gares de passagers du réseau de fer de la Bulgarie. On a appliqué les méthodes des composantes générales pour le groupement des facteurs et une analyse de cluster pour le classement des gares. D'après la méthode des composantes générales les facteurs sont classifiés en 4 groupes. D'après la méthode de l'analyse hiérarchique les gares sont classifiées en 6 groupes. Les méthodes de liaison moyenne entre le groupe et les intra-groupes de liaison et la distance de type Euclidienne au carré ont été comparées à la verification des résultats de l'analyse de cluster. Le groupement des gares est utilize pour déterminer les arrêts des cathégories des trains de passagers interurbains. Les groups de gares essentiels concernant le service des trains express sont premier, deuxième, troisième, quatrième et cinquième groupe. Le sixième groupe comprend les gares pour le service des TGV. La méthodologie peut être appliquée pour l'étude de toutes les gares et stations dans le réseau de chemin de fer.

1. INTRODUCTION

The development of the organisation of passenger transport by rail and transport services is dependent on the significance of the station in terms of volume of passenger traffic and also the potential of the settlement, which the station serves. The classification of passenger railway stations can be performed using one or more indicators, such as the scheme at the station plan and profile, number of tracks, the ability to transfer between tracks, the number of passengers departed for a period of time, the availability of commercial sites, the need for reconstruction and other indicators.

Each classification could serve for different purposes and it is important for the development of railway stations as an attractive centre for transportation. The classification can contribute to strategic planning, land use planning, the guidance of investments and management, and the quality of stations.

It can be summarised that the classification of railway stations is aimed at dealing with the following problems: transport – improving accessibility and connections at the station with other modes of transport; commercial – rehabilitation and development of stations as shopping centres, increasing their attractiveness; integrated – development of the areas around the stations as a place of transport and commercial services.

The aim of this research is to create a methodology for the classification of the railway passenger stations depending on the parameters characterising the potential of the location, infrastructure and transport volumes, which could be used as a basis for the development of transport services to different categories of passenger trains. This would allow developing and comparing the technological schemes for railways lines or complex railway network. Suburban trains and ordinary passenger trains usually serve all stations and stops along the route that they move. Therefore, the paper studies the stations that serve fast and express trains.

The object of the research is 98 railway stations in the Bulgarian railway network. The methodology of the study includes the following: determination of the indicators to examine the stations according to their characteristics; application of the data analysis methods for studying railway stations; and elaboration of the classification of the studied railway stations by using their pre-defined characteristics.

2. LITERATURE REVIEW

The classification of the railway stations is the subject of research of many authors. The purpose of these surveys is to be of service in the use of land and in sustainable development; transport and business development planning of the stations and their area. Both qualitative and quantitative criteria are used. The mathematical models which are applied are cluster analyses, regression models, ranking of the gravity of the factors in point scales, and identification by answers to questions.

The node-place model elaborated in [1] is the basis that is used by some authors to prepare the classification of railway stations. To determine the node value or the transport provision of a location, four criteria are analysed: the number of train connections departing from a station; the type of train connections present at a station; the proximity to the central business district by rail; and the number of bus connections departing from a station. To determine the place value (the quantity and diversity of human activities) of a station area, six criteria are analysed: the size of the population around the station; the characteristics of the nearby workforce; and the degree of multifunctionality. In [1, 2], 15 indicators were proposed to evaluate station's node and place functionality. In [4], the node place model has been used to gain insight into the development dynamics of 99 station areas in Tokyo. In [11], all Swiss railway stations are assessed in terms of node and place functions and eleven of these indicators are used to study 1684 railway stations in the Swiss railway network. Some of these indicators are: number of trains departing from the station; number of buses and trams departing from the station; number of stations reachable within 20 min.; bicycle access; number of residents within 700 m; and the number of workers per economics sector. A two-step cluster analysis is applied for the model. The number of clusters is determined on the basis of the Bayesian Information Criterion. Five clusters (smallest station, small stations, and two clusters of mid-size stations, large to very large

stations) are formed in the research. The purpose of this classification was to determine the development of stations. In [15], cluster analysis is applied – algorithm of Ward and Squared Euclidean distance measure of the 1700 Swiss railway stations as a tool for strategic planning, investments and quality of stations. Included are the factors describing the location of transportation infrastructures, properties of the catchment area, and properties of the public transport services at the station. The node-place model has also been applied in the Netherlands [3, 9]. The Dutch railways used this model to coordinate the strategies within their business units (passengers, station services, and real estate).

In order to make a classification of the railway stations and stops in the Czech railway network in [7], researchers used four groups of input parameters – position of the station within the network; position of the station as a public transport change node; settlement and a position of the station relatively to this settlement; and attractiveness of the station's surroundings. The stations were separated into five groups (A, B, C, D, and E) by this model. The purpose of the classification was to specify the conclusions for station equipment.

The model for categorisation of railway stations in the German railway network is presented in [5, 6, 10]. There, 5400 railway stations were observed and they were divided into seven categories. These categories are also a part of a price rating methodology named "SPS 11 – Stationspreissystem", which is used for setting the fee level for an external use of the railway stations. The model includes six criteria—number of platform edges; length of platforms; number of passengers per day; number of stopping trains per day; personnel present; and barrier-free platform access. The elaborated classification serves to determine the importance and general equipment of the station, which is a part of a pre-determined category.

The Rete Ferroviaria Italiana (RFI) classifies railway stations into Platinum, Gold, Silver and Bronze categories [10, 12]. The four parameters are used to make the classification: dimensions of the plant, which is the set of areas and surfaces accessible and attended by the travellers; the attendance, or the number of travellers and ordinary visitors who use the rail system daily; the ability of a rail system to connect with other public transport systems; the level of the commercial, or the quality of passenger service offered by the rail traffic.

A strategy for the classification of train station based on the sample data in North China and the use of unsupervised neural networks classification algorithms is presented in [8]. The investigated factors are: the passenger flow, train's origin and number, population quantity, and regional administrative level. The application of the Self-Organising Map classifies the train station into three clusters showing distinguishable levels.

It can be summarised that the classification of railway stations is aimed to deal with the following problems: transport – improving accessibility and connections at the station with other modes of transport; commercial – rehabilitation and development of stations as shopping centres, increasing their attractiveness; integrated – development of the areas around the stations as a place to transport and commercial services.

Most of these studies indicate that the cluster analysis may be successfully used for examining the railway stations. In the papers, how the classification of passenger railway stations can be used in the planning of the stops of different categories of trains has not been studied.

3. METHODOLOGICAL APPROACH

3.1. Factors for classification

The railway station is evaluated according to the settlement it serves. The factors which serve to characterise the stations according to the stops of the fast and express trains are chosen based on the parameters characterising the potential of the location, infrastructure and transport volumes. In this research the accessibility of railway stations in the city (distance to city centre), the development of the areas around the stations as a place of transport and commercial services are not considered as

factors for the classification of the stations in planning the stops of different categories of trains. The location of the station in the city depends on the design of the railways and urban development.

Four groups of factors associated with the resource potential of the location and infrastructure of the railway station are examined:

1. Potential of the town where the station is located. This includes demographics as well as territorial and socio-economic indicators.
   - H_1 – Population density (number of inhabitants / km^2).
   - H_2 – Area of the town, km^2.
   - H_3 – Number of learners. This factor determines the significance in terms of educational potential and shows the potential number of clients who use discounts when traveling by rail.
   - H_4 – Number of working-age population (over 18 years). This factor provides information about the potential population that needs jobs and can use railway transport services to and from work.
   - H_5 – Number of senior citizens (over 63 years). This factor indicates the potential number of clients using the discounts when travelling.
   - H_6 – Number of schools. This factor determines the significance in terms of educational potential.
   - H_7 – Number of universities. This factor also shows the importance in terms of educational potential.
   - H_8 – Average wage BGN/month. This indicator can assess the economic level of the town. Settlements with high values of this parameter can be centres of gravity for work trips.

2. Importance of the town where the station is located as an administrative, economic, industrial and cultural centre.
   - C_1 – Regional centre. The region is a large territorial unit and covers several areas. For example, Bulgaria is separated into six planning regions composed of 4 or 5 districts: the South-West, the South-Central, the South-East, the North-East, the North-Central and the North-West regions; each of them has a regional economic centre. These cities have a population of over 100 thousand people. $C_1 = 1$ if the city is a regional center; $C_1 = 0$, otherwise.
   - C_2 – District centre. The area is an administrative-territorial unit in the contemporary administrative-territorial division. For example, in Bulgaria there are 28 areas. These indicators can take values 0 or 1. $C_2 = 1$ if the city is a regional centre; $C_2 = 0$ otherwise.
   - C_3 – Participation of the town as an administrative and economic unit. $C_3 = 0$ if the city does not fulfil the function of the municipal, district or regional centre; $C_3 = 1$ otherwise.

3. Infrastructure factors
   - I_1 – Number of foster-starting tracks cash. This factor indicates the technical capabilities of the station to send and receive trains and is connected to its traffic option.
   - The importance of the station regarding its location on the rail network is assessed by indicators I_2 and I_3.
   - I_2 – A junction station that provides a link between the major railway lines (big junctions). These junctions also offer connections at the station with other modes of transport. This indicator shows the potential for the transfer of passengers between the railway lines. $I_2 = 1$ if the station performs those functions; $I_2 = 0$, otherwise.
   - I_3 – The junction station providing connections between the main and secondary railway lines (small junctions). This indicator shows the potential for the transfer of passengers and links between long-distance and suburban railway traffic. $I_3 = 1$ if the station performs those functions; $I_3 = 0$, otherwise.

4. Characteristics of passengers (number of departing passengers)
   - P_1 – Number of passengers for the month departed by fast and express train from the station of the railway network. This factor takes into account the workload of the station of intercity passengers, pass./month.

- P_2 – Number of passengers departed for the month with all categories of passenger trains in the station of the railway network number. This factor takes into account the overall workload of the station, pass./month.
- P_3 – Travelled passenger – kilometres with all categories of passenger trains, pass.km/ month. It is equal to the sum of the length (in km) of the journey travelled by each passenger departing from the station. This factor presents the measurement of passenger railway transport starting from the station.
- P_4 – Average distance travelled by passengers from the station, km. It is calculated by dividing the number of travelled passengers per kilometres by the number of departing passengers from the station. This factor indicates the average length of day trips from the station.

The data for determining the value of factors are obtained from National Statistical Institute of Bulgaria, the National Railway Infrastructure Company and the "BDZ Passengers services" Ltd. The database covers a five-year period from 2010 to 2014. The factors are defined as average values for this period.

3.2. Principal Component Analysis

The multivariate statistical method of the Principal Component Analysis (PCA) reduces the number of variables while retaining as much information about the original data matrix ($m×m$) by finding p new variables; where p is less than m, then the original variables can be written as a new data matrix ($p×m$), [14]. In the research the data matrix includes all factors determined above.

The new variables can account for the majority of information values of the original data when they are kept mutually orthogonal and uncorrelated. Every principal component is a linear function of original variables, and so it is often possible to ascribe meaning to what original variables represent. The principal components do not correlate with each other.

The verification of the derived factor model is performed using the following indicators:

- The used variable must correlate with each other and with the dependent variables, i.e. the bivariate correlation coefficients must be high. Otherwise, the factor analysis is not recommended.
- The model is verified by a Kaizer-Mayer-Olkin test (KMO test) and Bartlett's test of sphericity. The value of KMO test must be bigger than 0,5. The values of KMO test higher than 0,9 indicate a marvellous degree of common variance; higher than 0,8 are meritorious; higher than 0,7 are middling; higher than 0,6 are mediocre; higher than 0,5 are miserable; from 0 to 0,49 should not be factored. Bartlett's test of sphericity verifies whether the correlation matrix is an identity matrix, which would indicate that the factor model is inappropriate. The determinant of the matrix of data is converted to chi-square statistics and tested for significance. It is necessary for its associated probability to be less than 0,05.
- The grouping of the factors in the rotational matrix must be precise, i.e. a given variable can be a part of only one factor (it must have a strong correlation with it and a weak correlation with the other factors).

The main steps of a principal component analysis are:
- Normalisation of the original data set. Through normalisation of the original data, the influences over the research, which are brought by different dimensions and order of magnitude, are able to be removed.
- Finding the correlation matrix. Among the matrix is the correlated coefficient between x_i and x_j. Verification of the statistical significance of the correlation matrix can be used to Bartlett's test of sphericity.
- Finding the eigenvectors and eigenvalues of the correlation matrix.
- Computing the cumulative variance contribution rate of k principal components.
- Determining the number of principal components. Taking the several forward components with their cumulative variance contribution rate getting to 80% or 85%.
- Computing comprehensive factors scores of each case.

After choosing the number of factors to retain, it is necessary to spread variability more evenly among factors. The factor analysis for rotating the factors is used for this. The study has made use of the Varimax method (maximised squared loading variance across variables) of orthogonal rotation as a technique of rotation. The rotation is a stage of the factor analysis in which the factor matrices with small differences between the factor coefficients is transformed into a relatively simpler one. The obtained coefficients differ to a greater degree.

3.3. Cluster analysis

The cluster analysis is a suitable method for the classification of the examined railway stations into groups by using different factors. It is a multi-measurable statistical analysis for a classification of units into groups, preliminarily unknown, based on numerous characteristics in relation to these units [13, 14]. In is necessary for the number of examined factors to be greater than 2. In the research, the factors, obtained by Principal Component Analysis, are applied.

The Statistics theory suggests different methods of clusterisation. A method for hierarchical clustering has been used in the study. The main advantage of this method is that the determination of a unit into a specific cluster is definitive. Hierarchical clustering is performed by the agglomerative method. For the distance-type measures, the Squared Euclidean distance and Euclidean distance [13] were chosen.

The dispersion analysis could be used for an approximate evaluation of the clusterisation results as well as for determining the roles of each variable used for the clusters' establishment. The determination of the statistical importance of different factors is done by using the Fisher's criterion.

$$F \geq F_T \tag{1}$$

where: F is the empirical value of the criterion resulted from the dispersion analysis, F_T is the theoretical value when the level of risk $\alpha = 0,05$ and the number of degrees of freedom, $k_1 = n-1$; $k_2 = m-n$; m is the number of observations (number of railway stations), n is the number of examined factors (number of principal components). The theoretical criterion F_T is determined by using the tables for F distribution.

On the one hand, the Fisher' criterion's evaluation determines which factors are significant for the study. On the other, it does not dismiss those other factors which are used for clusterisation but do not satisfy the condition (1). The F tests should be used only for descriptive purposes because the clusters have been chosen to maximise the differences between cases in different clusters.

4. APPLICATION. A STUDY FOR BULGARIAN RAILWAY NATWORK

There are 307 passenger stations in Bulgaria's railway network. The research was confirmed for 98 of the stations, located on the major railway lines. They are serviced by all the categories of passenger trains. The research did not include stations which are serviced only by standard passenger or suburban trains. In the present investigation, all of the aforementioned 18 factors were included. The research was conducted in the following stages: determining the values of the factors for each of the stations, conducting a factor analysis, conducting a cluster analysis.

5. APPLICATION - A STUDY FOR THE BULGARIAN RAILWAY NETWORK

There are 307 passenger stations in Bulgaria's railway network. The research was confirmed for 98 of the stations, located on the major railway lines. They are serviced by all the categories of passenger trains. The research did not include stations which are serviced only by standard passenger or suburban trains. In the present investigation, all of the aforementioned 18 factors were included. The research was conducted in the following stages: determining the values of the factors for each of the stations, conducting a factor analysis, and conducting a cluster analysis.

5.1. Classification of criteria using Principal Component Analysis

The SPSS (Statistical Package for Social Science) software has been used for carrying out the study with a Principal Component Analysis and Cluster Analysis. Before proceeding to the extraction of factors, the data regarding the requirements for the application of factor analysis must be evaluated. The first stage is to investigate the correlation matrix. The determinant of the correlation matrix is shown at the foot of the table below. To carry out an analysis, it is necessary that the determinant is greater than 0, Tab. 1.

Fig. 1 shows the values of communality for the observed data. The communality shows the total influence on a single observed variable from all the factors associated with it. It is equal to the sum of all the squared factor loadings for all the factors related to the observed variable and this value is the same as coefficients of determination (R^2) in multiple regression. The value ranges from zero to 1 where 1 indicates that the variable can be fully defined by the factors and has no uniqueness. The results show that low coefficients of determination have a part of the settlement as an administrative entity (C3) and average distance (P4).

Table 1

Correlation matrix

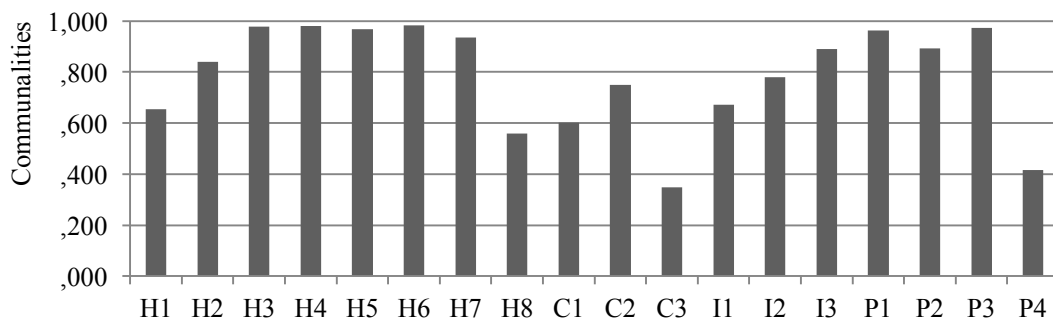|	H1	H2	H3	H4	H5	H6	H7	H8	C1	C2	C3	I1	I2	I3	P1	P2	P3	P4
H1	1,00	,06	,41	,36	,38	,44	,46	-,10	,42	,64	,21	,36	,22	-,06	,37	,47	,39	,17
H2	,06	1,00	,89	,91	,90	,86	,79	,43	,54	,31	,05	,43	,22	-,09	,87	,74	,86	,27
H3	,41	,89	1,00	,99	1,00	,99	,95	,35	,74	,51	,12	,53	,30	-,08	,96	,89	,96	,29
H4	,36	,91	,99	1,00	,99	,99	,95	,38	,71	,45	,10	,53	,30	-,07	,97	,89	,97	,27
H5	,38	,90	1,00	,99	1,00	,98	,94	,36	,72	,49	,11	,51	,31	-,08	,95	,89	,95	,26
H6	,44	,86	,99	,99	,98	1,00	,97	,34	,73	,51	,12	,54	,31	-,08	,95	,93	,96	,26
H7	,46	,79	,95	,95	,94	,97	1,00	,35	,74	,43	,08	,50	,32	-,07	,94	,94	,96	,22
H8	-,10	,43	,35	,38	,36	,34	,35	1,00	,24	-,07	-,06	,11	,20	-,12	,38	,35	,38	-,08
C1	,42	,54	,74	,71	,72	,73	,74	,24	1,00	,42	,08	,47	,30	-,07	,68	,68	,73	,37
C2	,64	,31	,51	,45	,49	,51	,43	-,07	,42	1,00	,19	,28	,05	-,17	,40	,40	,40	,30
C3	,21	,05	,12	,10	,11	,12	,08	-,06	,08	,19	1,00	,12	,14	,10	,10	,10	,10	,22
I1	,36	,43	,53	,53	,51	,54	,50	,11	,47	,28	,12	1,00	,46	,21	,58	,56	,58	,32
I2	,22	,22	,30	,30	,31	,31	,32	,20	,30	,05	,14	,46	1,00	-,12	,35	,38	,34	,11
I3	-,06	-,09	-,08	-,07	-,08	-,08	-,07	-,12	-,07	-,17	,10	,21	-,12	1,00	-,03	,00	-,03	-,14
P1	,37	,87	,96	,97	,95	,95	,94	,38	,68	,40	,10	,58	,35	-,03	1,00	,93	,99	,26
P2	,47	,74	,89	,89	,89	,93	,94	,35	,68	,40	,10	,56	,38	,00	,93	1,00	,93	,11
P3	,39	,86	,96	,97	,95	,96	,96	,38	,73	,40	,10	,58	,34	-,03	,99	,93	1,00	,30
P4	,17	,27	,29	,27	,26	,26	,22	-,08	,37	,30	,22	,32	,11	-,14	,26	,11	,30	1,00
Determinant=3,70E-016

Fig. 1. Coefficients of determination of the factors
Fig. 1. Coefficient de determination des facteurs

In order to know whether the indices are suitable for processing factor analysis, the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity are examined. Table 2 shows the value of the KMO.

The value is 0,846>0,5 and means that conducting factor analysis is good enough. From the same table, we can see that the Bartlett's test of sphericity is significant. The associated probability is less than 0,05 (i.e. the significance level is small enough) which indicates that the data are multivariate normal and the correlation matrix is statistically significant and acceptable for factor analysis.

Table 3 shows all the factors extractable from the analysis along with their eigenvalues, the percent of variance attributable to each factor, and the cumulative variance of the factor and the previous factors. Notice that the first factor accounts for 56,039% of the variance, the second for 9,877%, the third for 6,751% and the fourth for 6,113%. All the remaining factors are not significant.

Table 4 shows the component matrix before and after rotation. The results of component matrix show that the parameter Z_2 cannot determine exact affiliation to component 1 or component 2. This is because the values are similar.

Table 2

Values of Kaiser-Meyer-Olkin (KMO) test
and Bartlett's test of sphericity

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,846
Bartlett's Test of Sphericity	Approx. Chi-Square	3203,799
	df	153
	Sig.	,000

Table 3

Eigenvalues of explained variance

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10,087	56,039	56,039	10,087	56,039	56,039	9,260	51,447	51,447
2	1,778	9,877	65,915	1,778	9,877	65,915	2,096	11,645	63,092
3	1,215	6,751	72,666	1,215	6,751	72,666	1,654	9,189	72,281
4	1,100	6,113	78,779	1,100	6,113	78,779	1,170	6,498	78,779

Table 4

Component matrix and rotated component matrix

Component Matrix[a]

	Component			
	1	2	3	4
H1	,466	,652	-,068	,086
H2	,858	-,311	-,086	-,031
H3	,983	-,051	-,085	,044
H4	,980	-,115	-,062	,047
H5	,976	-,084	-,075	,044
H6	,986	-,036	-,070	,062
H7	,961	-,064	-,052	,077
H8	,367	-,604	,029	-,242
Z1	,744	,209	-,072	-,012
Z2	,516	,596	-,344	,099
Z3	,141	,445	,357	-,037
I1	,610	,198	,501	-,103
I2	,368	,028	,579	-,555
I3	-,072	-,037	,576	,743
P1	,972	-,115	,041	,050
P2	,930	-,074	,081	,121
P3	,980	-,096	,026	,041
P4	,320	,442	-,020	-,342

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
H1	,292	,739	,151	-,006
H2	,901	-,100	,070	-,121
H3	,962	,184	,118	-,070
H4	,974	,118	,120	-,054
H5	,963	,148	,117	-,063
H6	,962	,198	,124	-,047
H7	,945	,166	,119	-,023
H8	,470	-,540	,098	-,195
Z1	,658	,365	,163	-,098
Z2	,384	,756	-,075	-,156
Z3	-,024	,367	,432	,159
I1	,461	,192	,629	,163
I2	,207	-,115	,838	-,146
I3	-,017	-,038	,036	,942
P1	,956	,096	,197	,008
P2	,911	,130	,191	,091
P3	,959	,116	,196	-,008
P4	,144	,426	,336	-,317

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

The parameter I2 has small differences between the factor coefficients. Therefore, adherence to a component is not enough definitely. It is, therefore, a necessary rotation of the components. In this research the Varimax method for rotation is applied. The rotated component matrix shows the final forming of components. Table 5 shows the individual factor score. As we have four factors, we have four-factor equations, which are compiled using the values from the component/factor score coefficient matrix.

The first component includes the parameters of three of the investigated groups: area of the town in km^2; number of learners; number of the working-age population; number of senior citizens; number of schools; number of universities. The second component includes the following: population density (number of inhabitants/km^2); average wage BGN/month; district centre; and average distance travelled by passengers from the station in km. The third component includes the parameters: participation of the town as an administrative and economic unit; number of foster-starting tracks cash; and junction station that provides a link between the major railway lines. The fourth component accounts for only a parameter junction station providing a connection between the main and the secondary railway lines.

Table 5

Factor equations and Component score coefficient matrix

Factor equations	Component score coefficient matrix			

Factor equations:

$$F_1 = -0,035H_1 + 0,13H_2 + 0,113H_3 + 0,12H_4 + 0,116H_5$$
$$+ 0,112H_6 + 0,114H_7 + 0,091H_8$$
$$+ 0,044C_1 + 0,003C_2 - 0,087C_3 - 0,027I_1$$
$$- 0,088I_2 + 0,042I_3 + 0,111P_1 + 0,107P_2$$
$$+ 0,109P_3 - 0,075P_4$$

$$F_2 = 0,38H_1 - 0,137H_2 + 0,015H_3 - 0,022H_4 - 0,005H_5$$
$$+ 0,023H_6 + 0,007H_7 - 0,356H_8$$
$$+ 0,136C_1 + 0,4C_2 + 0,171C_3 + 0,015I_1$$
$$- 0,165I_2 - 0,04P_1 - 0,014P_2 - 0,028P_3$$
$$+ 0,189P_4$$

$$F_3 = 0,007H_1 - 0,061H_2 - 0,061H_3 - 0,055H_4$$
$$- 0,059H_5 - 0,058H_6 - 0,057H_7$$
$$+ 0,067H_8 + 0,002C_1 - 0,184C_2$$
$$+ 0,308C_3 + 0,413I_1 + 0,654I_2 + 0,009I_3$$
$$+ 0,01P_1 + 0,005P_2 + 0,006P_3 + 0,214P_4$$

$$F_4 = 0,024H_1 - 0,066H_2 - 0,013H_3 + 0,001H_4$$
$$- 0,007H_5 + 0,008H_6 + 0,028H_7$$
$$- 0,164H_8 - 0,049C_1 - 0,097C_2$$
$$+ 0,136C_3 + 0,154I_1 - 0,143I_2 + 0,824I_3$$
$$+ 0,051P_1 + 0,123P_2 + 0,038P_3 - 0,269P_4$$

Component score coefficient matrix:

	Component			
	1	2	3	4
H1	-,035	,380	,007	,024
H2	,130	-,137	-,061	-,066
H3	,113	,015	-,061	-,013
H4	,120	-,022	-,055	,001
H5	,116	-,005	-,059	-,007
H6	,112	,023	-,058	,008
H7	,114	,007	-,057	,028
H8	,091	-,356	,067	-,164
C1	,044	,136	,002	-,049
C2	,003	,400	-,184	-,097
C3	-,087	,171	,308	,136
I1	-,027	,015	,413	,154
I2	-,088	-,165	,654	-,143
I3	,042	,000	,009	,824
P1	,111	-,040	,010	,051
P2	,107	-,014	,005	,123
P3	,109	-,028	,006	,038
P4	-,075	,189	,214	-,269

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

5.2. Classification of railway passenger station using Cluster Analysis

The Cluster Analysis is applied to classify the investigated railway stations. The four components obtained by Principal Component Analysis are used as a criterion. All of the 98 stations were tested with them. Table 6 shows the value of F criterion for the examined indicators, which are defined by the Principal Component Analysis. The theoretical value of F criterion is defined by using standard tables and is $F_T = 2,68$; $k_1 = 98-4 = 94$, $k_2 = 4-1 = 3$, level of risk $\alpha = 0,05$.

Table 6

Values of F criterion

Indicator	F1	F2	F3	F4
F criterion	309,327	53,218	47,418	152,089

From the results, which are shown in Table 6, the importance of each one of the factors can be determined when a classification has been conducted.

In the study, the method of hierarchical clustering is applied. Two agglomerative methods are applied in which successive mergers of units in groups are carried out − Average linkage between groups (ALBG) and Within-groups linkage (WG). Each of the methods has been studied with two distance-type measures − Euclidean distance (E) and Squared Euclidean distance (SqE). Fig. 2a and Fig. 2b present dendrograms of cluster tree of railway stations according to Euclidean distance and Squared Euclidean distance. Fig. 3 shows the number of clusters for each variant. The results show a similar disposition of surveyed stations in clusters. This gives a reason to assume that the results are verified.

6. CONCLUSIONS

The study has shown the following results:

• The factors for classifications of railway passenger stations have been defined. The main groups of these factors are the potential of the town, the importance of the town, infrastructural factors, and the characteristics of passengers. In the research are defined 18 factors.

• The Principal Component Analysis has been applied to classify the factors. They have been separated into four components.

• A cluster analysis has been used for the classification of the railway passenger stations. The classification has been conducted by using four components defined by a Principal Component Analysis.

• The methods Average linkage between group and the Within-groups linkage, and the distance-type measures Euclidean distance and the Squared Euclidean distance applied to these methods were compared to verify the results of the cluster analysis.

• The application of the Cluster analysis and The Principal Component Analysis allows us to evaluate the railway stations.

• A classification of railway passenger station was proposed by the method Average linkage between group and distance-type measures Euclidean distance. The station was divided into six groups. The results indicate that the railway passenger stations can be classified into six groups:
   o   Group 1: Includes the biggest passenger railway station—Sofia Central Station.
   o   Group 2: Includes 6 stations, which are small junction stations.
   o   Group 3: Includes 4 stations located in big cities with more than 100 thousand inhabitants.
   o   Group 4: Includes 12 stations. These are the large junction stations.
   o   Group 5: Includes 16 stations. These are stations located in cities which are regional centres.
   o   Group 6: Includes 59 stations. This group is subdivided into two more groups. The first cluster (6a) includes cities, which are not municipal centres (10 stations), and the second (6b) includes stations in cities which are municipal centres.

• This classification can be used in determining the stops of the different categories of passenger trains. It can be concluded that the main clusters of stations to be serviced by express trains should be Groups 1, 2, 3, 4 and 5. Group 6 is a cluster to be serviced by fast passenger trains. Different versions of servicing the cities with intercity trains can be developed using the same classification.

• The methodology can be applied to study all stations and stops on the rail network.
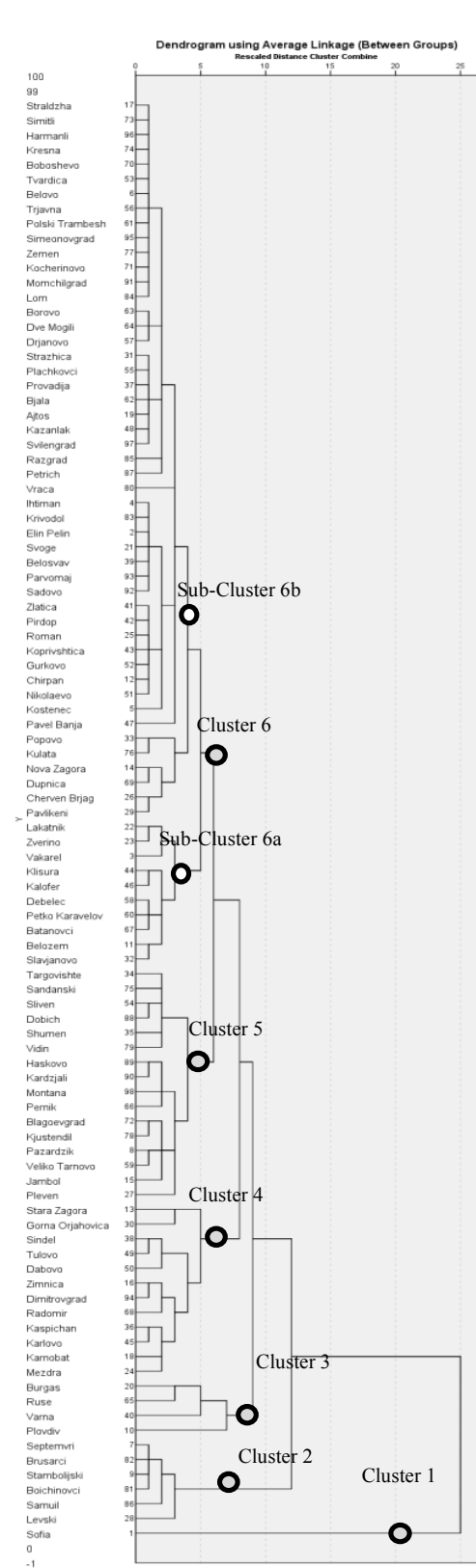
Acknowledgement

Fig. 2a. Dendrogram of cluster tree of railway stations
         according to Euclidean distance
Fig. 2a. Dendrogramme d'arbre de cluster des gares
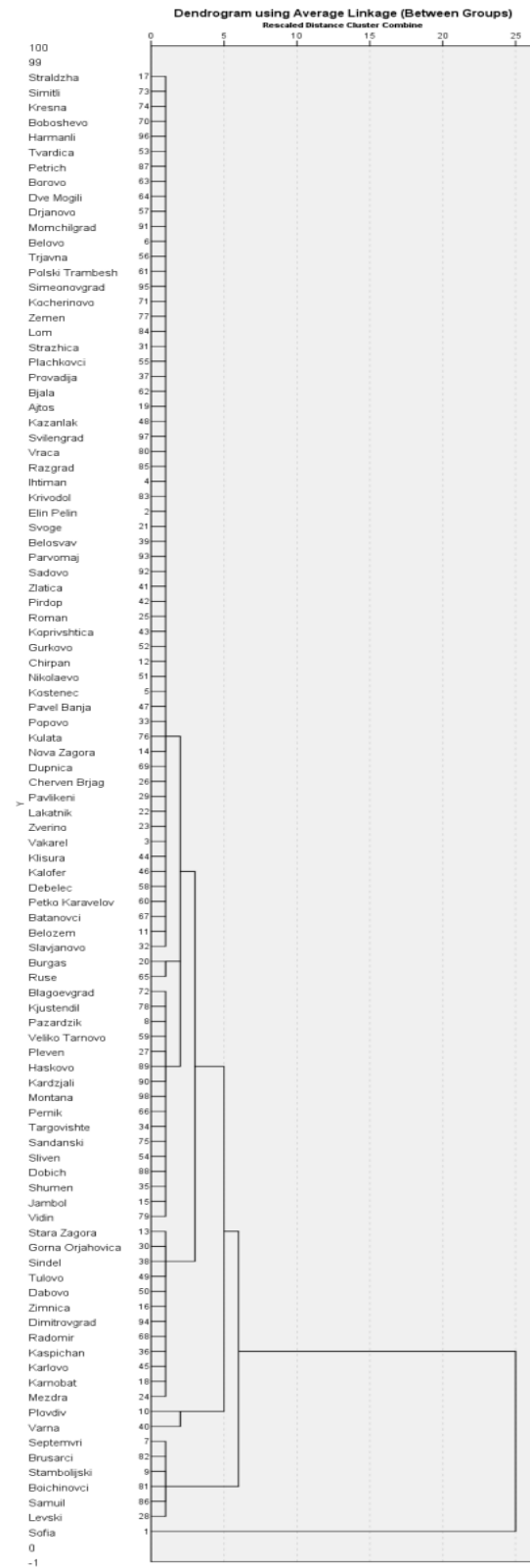   selon la distance Euclidienne

Fig. 2b. Dendrogram of cluster tree of railway stations
         according to Squared Euclidean distance
Fig. 2b. Dendrogramme d'arbre de cluster des gares
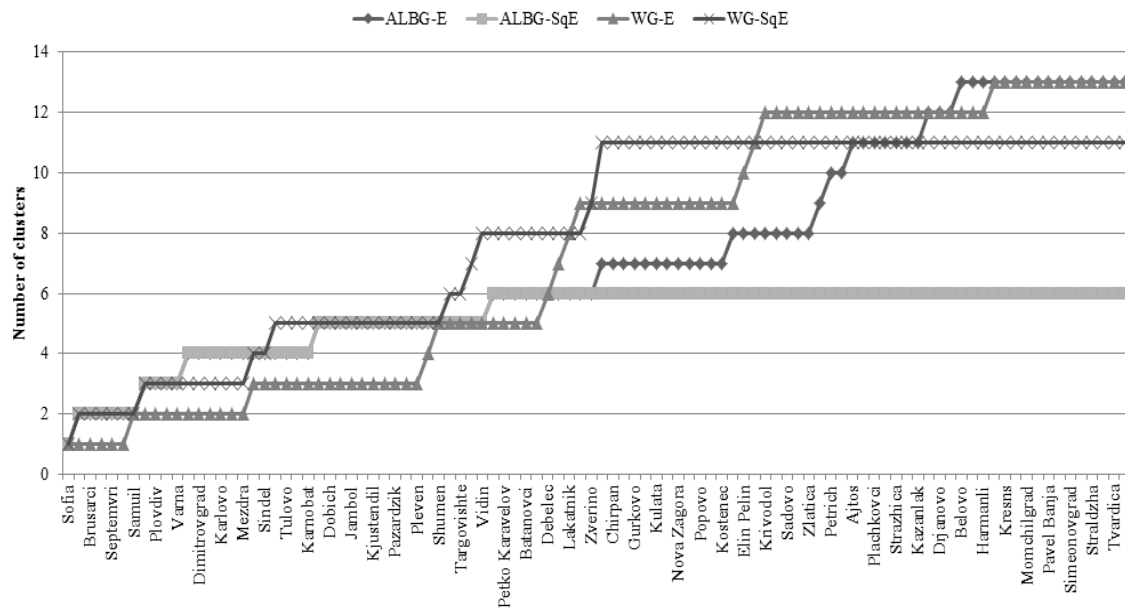   selon carré de la distance Euclidienne

Fig. 3. Comparison of the variants of clustering
Fig. 3. Comparaison des variantes de la typologie (la théorie de Cluster)

References

1.   Bertolini, L. Nodes and places complexities of railway station redevelopment. *European Planning Studies*. 1996. Vol. 4. No. 3. P. 331-34.
2.   Bertolini, L. Station area redevelopment in five European countries. An international respective on a complex planning challenge. *International Planning Studies*. 1998. Vol. 3. No. 2. P. 163-184.
3.   Bertolini, L. Spatial development patterns and public transport: the application of an analytical model in the Netherlands. *Planning, Practice and Research*. 1999. No. 14. P. 199–210.
4.   Chorus, P. & Bertolini, L. An application of the node- place model to explore the spatial development dynamics of station areas in Tokyo. *The Journal of Transport and Land Use*. 2011. Vol. 4. No. 1. P. 45-48.
5.   *Das Stationspreissystem SPS 11 - Gültig ab 01.01.2013*. Berlin: DB Station&Service AG. 2012. Available at:
     http://www.deutschebahn.com/file/3047500/data/stationspreissystem_2011_gueltig_ab_2013.pdf
     [In German: *The station pricing system SPS 11 - Valid from 01.01.2013*. Berlin: DB Station & Service AG.]
6.   *Die sieben Bahnhofskategorien*. Berlin: DB Station&Service AG. 2013. Available at: http://www.deutschebahn.com/de/geschaefte/infrastruktur/bahnhof/bahnhofs_kategorien.html.
     [In German: *The seven categories of railway station*. Berlin: DB Station & Service AG.]
7.   Havlena, O. & Jacura, M. et al. Parameters of passenger facilities according to railway station characteristics. *Transport Problems*. 2014. Vol. 9. No. 4. P. 97-104.
8.   Hai, X. & Zhao, C. & Jiang, X. Train Station Classification for Passenger Dedicated Line. *International Journal of Advancements in Computing Technology (IJACT)*. 2012. Vol. 4. No. 15. P. 328-335.
9.   Peek, G. & Bertolini, L. Gaining insight in development potential of station areas: A decade of Node-Place modelling in the Netherlands. *Planning, Practice and Research*. 2006. Vol. 21. No. 4. P. 443-462.
10. Pucci, P. Classifying railway stations. *Workshop on Infrastructures planning and design*. Available at: http://www.workshop-infra.polimi.it/lezioni/documenti/cassi.pdf

11. Reusser, D. & Loukopoulos, P. & Scholz, R. Classifying railway stations for sustainable transitions – balancing node and place functions. *Journal of transport geography*. 2008. No. 16. P. 191-202.

12. *Sustainability Report 2014*. Ferroviaria Italiana. Strategy and Planning Head Office. 2015. Available at:
http://www.fsitaliane.it/cms-file/allegati/fsitaliane_en/2014_Sustainability_Report.pdf

13. Srivastava, T. & Shailaja, R. *Statistics for Management*. McCraw-Hill Publishing. 2008. 266 p.

14. Tan, P. & Steinbach, M. & Kumar, U. *Introduction to Data mining*. Minnesota: University of Minnesota. 2006. 769 p.

15. Zemp, S. & Stauffacher, M. et al. Classifying railway stations for strategic transport and land use planning: Context matters. *Journal of transport geography*. 2011. No. 19. P. 670-67.