



Journal Homepage: - [www.journalijar.com](http://www.journalijar.com)

## INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

Article DOI: 10.21474/IJAR01/10576

DOI URL: <http://dx.doi.org/10.21474/IJAR01/10576>



### RESEARCH ARTICLE

#### FACTOR ANALYSIS IN SEISMOLOGY

R.Z. Burtiev, V.G. Alcaz, S.V. Troian and V.Yu. Kardanets

Institute of Geology and Seismology of Moldova, Chisinau.

#### Manuscript Info

##### Manuscript History

Received: 22 December 2019

Final Accepted: 25 January 2020

Published: February 2020

##### Key words:-

PSHA, Factor Analysis, SPSS

#### Abstract

Factor analysis allows us to explore the structure of the relationship of variables, where each group of variables will be determined by the factor according to which these variables have maximum loads. The result of factor analysis is a transition from a set of input parameters to a smaller number of variables, called factors. The factor is interpreted as a hidden variable - the cause of joint variability (interrelation) of several initial parameters. A method for calculating of the attenuation function values is proposed. On the basis of the obtained values, the seismic hazard is calculated as the probability that for a fixed time  $t$  at points of the earth's surface  $Q(\varphi, \psi)$   $n$  seismic shocks occur, of which  $m$  will have intensity  $I^k$  of the MSK-64 scale. The method is used to study the seismicity of the Foci Vrancea and earthquakes of Romania. Seismic processes are complex and diverse, since their formation is due to complex, diverse geological and geophysical processes occurring in the bowels of the Earth, and is characterized by many different parameters, and the results of observations on them are represented as multidimensional random variables. In the study of such multiparameter processes, the question arises: is it possible to discard some of the parameters, or replace them with a smaller number of some functions of them, while retaining all the information? To solve this problem is a factor analysis [Bakhtin, et al., 2007]. Factor analysis is based on determining the minimum number of factors that make up the largest share in the variance of the data [Harman, 1972]. In the study of the complex nature of seismicity, factor analysis helps to better understand the essence of seismic processes, since the interdependence between seismic parameters must be due to relationships between parameters, the identification of which is the task of factor analysis.

Copy Right, IJAR, 2020,. All rights reserved.

#### Introduction:-

Seismology is mostly an empirical science, since statements, like models, are based on facts and observations. The goal of the empirical sciences is to cover the abstract model of the results of real-world observations. Presentation of the picture of future behavior and determination of the future form of the model is the main task of empirical sciences. In practical applications, the best is considered the model that is the most general and simplified description of observations, and has a high forecast potential. In order to draw conclusions about the general, real relationship in a set of observations with some (most) probability, methods of probability theory are used, since it is impossible to cover the whole set of observations with a model. The purpose of stochastic modeling is to recognize

Corresponding Author:- R. Z. Burtiev

Address:- Institute of Geology and Seismology of Moldova, Chisinau.

the true structure of the system from the observed data. The array of seismic data is analyzed, consisting of values of **17 parameters characterizing the seismicity of the Vrancea source:**

MSK - intensity at the epicenter;

Mw - moment magnitude;

R- is the hypocentral distance;

Az - azimuth to the most distant point;

NP<sub>1stk</sub> - the direction of the nodal plane 1;

NP<sub>1dp</sub> is the angle of incidence of the nodal plane 1;

NP<sub>1slip</sub> - offset of the nodal plane 1;

NP<sub>2stk</sub> - the direction of the nodal plane 2;

NP<sub>2dp</sub> is the angle of incidence of the nodal plane 2;

NP<sub>2slip</sub> - shift of the nodal plane 2;

P<sub>az</sub> - azimuth of the compression axis;

P<sub>pl</sub> is the angle of the compression axis;

B<sub>az</sub> is the azimuth of the neutral axis

B<sub>pl</sub> is the angle of the neutral axis;

T<sub>az</sub> - azimuth of the axis of stretching;

T<sub>pl</sub> is the angle of the stretching axis.

Modern models of the Earth's structure and theories explaining the occurrence of earthquakes are based on indirect data, mainly on seismic observations. The main purpose of geophysical research is to solve the inverse problem, i.e. determination of the structure of the environment from observations of the characteristics of physical fields. The attenuation of the intensity of seismic effects is one of the factors determining the quality of seismic hazard analysis. The characteristics of the seismic impact of each earthquake are determined by its characteristics such as tectonics (tectonic structure of the region), focal depth, mechanism, focal geometry, direction and course of the process of fracturing of rocks and other parameters. The picture of the macroseismic field is a reflection of the influence of all these factors, and local geological features on the manifestation of the seismic effect at the points of the day surface. One of the significant factors determining the quality of seismic hazard analysis is the seismicity model of the earthquake zone. Seismicity is the susceptibility of the Earth or individual territories to earthquakes, which is a reflection of the geological processes occurring in the earth's interior, and occur in areas of the Earth that are different in structure and nature of geological development. Seismicity is characterized by:

1. Earthquake frequency
2. Statistical distribution of the force of shocks (magnitudes),
3. Spatial distribution of foci,
4. Macroseismic observations of strong seismic events (seismic intensity, damage pattern).

In the study of seismicity, we are dealing with a seismic process occurring in time and space, and many of the tasks of seismology are associated with the ability to calculate the probabilities of some events associated with the seismic process and the determination of their probabilistic structure. Seismology is concerned with the study of earthquakes and related phenomena. The main task of seismology is to learn how to predict the strength, time and place of occurrence of earthquakes. The model, which covers various forms of dependence between events, is a real reflection of reality, but this leads to great difficulties in studying their probabilistic structure and statistical analysis. Therefore, a compromise decision is made - a model is selected that takes into account the dependencies sufficient for the adequacy of the model and is amenable to statistical analysis. Mathematical statistics, whose main task is to link the real world of data with the world of theoretical models, develops on the basis of probability theory and is used to determine the parameters of the model from sample data, forecast, hypothesis testing, and to some extent solves inverse problems. Mathematical statistics reveal the likelihood degree of some hypotheses regarding some phenomena. In addition, methods of mathematical statistics can help the researcher in drawing conclusions about mass phenomena, from observations on them, to identify the most significant factors and areas of research. Seismic processes occur and develop in time and space under the influence of the internal determinism of global tectonics. Uncertainties associated with the interweaving of the internal physical fields of the Earth and the gravitational force of celestial bodies, and their influence on global tectonics, introduce an element of randomness in seismicity models [Burtiev, 2017].

### **Methodology:-**

Seismic processes are complex and diverse, since their formation is due to complex, diverse geological and

geophysical processes occurring in the bowels of the Earth, and is characterized by many different parameters, and the results of observations on them are represented as multidimensional random variables. In the study of such multiparameter processes, the question arises: is it possible to discard some of the parameters, or replace them with a smaller number of some functions of them, while retaining all the information? To solve this problem is a factor analysis. Factor analysis is based on determining the minimum number of factors that make up the largest share in the variance of the data. In the study of the complex nature of seismicity, factor analysis helps to better understand the essence of seismic processes, since the interdependence between seismic parameters must be due to relationships between parameters, the identification of which is the task of factor analysis. Factor analysis allows you to explore the structure of the relationship of variables, where each group of variables will be determined by the factor according to which these variables have maximum loads. The result of factor analysis is a transition from a set of initial variables to a smaller number of new variable factors. The factor is interpreted as a hidden variable - the cause of joint variability (interrelation) of several source variables. Factor analysis solves the problem of reducing the number of variables with minimal loss of initial information. If we assume that the correlations between variables are explained by the influence of hidden causes - factors, then the main task of factor analysis is to analyze the correlations of the set of parameters

[Lindsay I Smith [Error! Hyperlink reference not valid.](#)  
[Error! Hyperlink reference not valid.](#)  
[Error! Hyperlink reference not valid.](#), 2009].

If the researcher is only interested in the structure of variables, this completes the factor analysis. Continuing factor analysis, we can calculate the values of the factors for further regression analysis. Most factor analysis methods are based on the principal component analysis (PCA), which converts a group of correlated source variables into another group of uncorrelated variables. PCA solves the problem of reducing the number of variables while maintaining the maximum proportion of the variance of observations, choosing only the main components that explain most of the variance.

The main stages of factor analysis: calculation of the correlation matrix; extraction of factors; selection and rotation factors; interpretation of factors; calculation of factor values; assessment of the quality of the model. The array of seismic data is analyzed that consists of the values of 16 parameters characterizing the mechanism and geometry of the Vrancea core: MSK-degree of intensity in epicenter of MSK-64 scale; Mw; R; Az; NP<sub>1stk</sub>; NP<sub>1dp</sub>; NP<sub>1slip</sub>; NP<sub>2stk</sub>; NP<sub>2dp</sub>; NP<sub>2slip</sub>; P<sub>az</sub>; P<sub>pl</sub>; B<sub>az</sub>; B<sub>pl</sub>; T<sub>az</sub>; T<sub>pl</sub>. It is assumed that the values of these parameters are due to hidden factors that cannot be directly observed, and the identification of which is the task of factor analysis. The task of factor analysis is to determine the minimum number of factors that contribute the most to data variance. The use of factor analysis begins with the calculation of descriptive statistics of an array of observed values over 16-dimensional random variables characterizing the seismicity mechanism and geometry of the Vrancea focus. Factor analysis can reveal relationships that determine the correlation interdependence between seismic parameters. It is assumed that the reason for the significant correlation between the parameters is the desired factors. The task of factor analysis is to determine the minimum number of factors that contribute the most to data variance. The values of the table (Table 1) show that the values of the statistical characteristics of the parameters have a large scatter, that is, their values are measured in different units that are incompatible with each other, so they should be brought to a single scale using standardization (1):

$$X_{ij}^s = \frac{X_{ij} - \bar{X}_j}{\sigma_j} \quad (1)$$

Where  $X_{ij}$  value of the  $j$ -th parameter in the  $i$ -th observation,  $\bar{X}_j$  mean and  $\sigma_j$  standard deviation of  $j$ -th parameter. According to the values of the table (Table 1), it can be seen that the values of the statistical characteristics of the parameters have a large scatter, and their values are measured in different units that are not comparable with each other, therefore they should be brought to a single scale using standardization. The first idea of the presence of dependent parameters can be obtained from the correlation matrix (Table 2), which characterizes the degree of correlation between the parameters of the original array. The higher the proportion of high correlations, the better the data are suitable for factor analysis. For example, the value of the correlation coefficient between the parameters

np<sub>1stk</sub> and np<sub>2stk</sub>, contains 0.726 (table 2). The value of the correlation coefficient between the paz parameter and np<sub>1dp</sub> is equal to 0.884, which indicates a rather high degree of dependence between the parameters and is the basis for their inclusion in one group.

**Table 1:-** Descriptive statistics.

Parameters	Mean	Std. Deviation	N
M <sub>w</sub>	6,8200	0,68015	9914
lnR	641,4406	776,26918	9914
Azim	176,8295	95,23605	9914
NP <sub>1stk</sub>	206,0295	61,68680	9914
NP <sub>1dp</sub>	53,7408	18,12259	9914
NP <sub>1slip</sub>	92,1033	29,92217	9914
NP <sub>2stk</sub>	84,7021	108,54354	9914
NP <sub>2dp</sub>	38,8398	19,16654	9914
NP <sub>2slip</sub>	73,3290	28,41213	9914
P <sub>az</sub>	209,8088	114,53092	9914
P <sub>pl</sub>	18,9177	10,19420	9914
B <sub>az</sub>	68,8550	45,65304	9914
B <sub>pl</sub>	12,1869	5,55001	9914
T <sub>az</sub>	190,5526	60,39683	9914
T <sub>pl</sub>	67,0351	9,75051	9914

Units are located along the main diagonal of the correlation matrix, as well as the covariance matrix, and they represent the variances of the used I-parameters; however, due to the standardization of the parameters, these variances become 1.

**Table 2:-** Correlation Matrix.

	MSK	MW	lnR	azim	np1stk	np1dp	np1slip	np2stk	np2dp	np2slip	paz	ppl	baz	bpl	taz	tpl
MSK	1,000	,366	-,458	,010	-,016	,114	,114	-,089	-,137	,093	,163	,146	-,261	-,191	-,186	-,037
MW	,366	1,000	-,704	,192	,430	,367	,559	-,513	-,370	,066	,582	-,043	-,713	-,003	-,270	,147
lnR	-,458	-,704	1,000	-,194	-,066	,029	-,334	,046	-,030	-,175	-,118	-,070	,554	-,016	-,116	,047
azim	,010	,192	-,194	1,000	,082	,082	,103	-,108	-,083	-,005	,120	-,008	-,174	,013	-,041	,035
np <sub>1stk</sub>	-,016	,430	-,066	,082	1,000	,543	,665	-,820	-,488	-,329	,580	-,469	-,506	,379	,034	,278
np <sub>1dp</sub>	,114	,367	,029	,082	,543	1,000	,223	-,830	-,993	-,336	,884	,277	-,535	-,315	-,623	-,101
np <sub>1slip</sub>	,114	,559	-,334	,103	,665	,223	1,000	-,447	-,229	,357	,477	-,672	-,620	,092	-,040	,674
np <sub>2stk</sub>	-,089	-,513	,046	-,108	-,820	-,830	-,447	1,000	,796	,323	-,862	,065	,599	-,098	,432	-,082
np <sub>2dp</sub>	-,137	-,370	-,030	-,083	-,488	-,993	-,229	,796	1,000	,268	-,895	-,283	,545	,397	,677	,056
np <sub>2slip</sub>	,093	,066	-,175	-,005	-,329	-,336	,357	,323	,268	1,000	-,093	-,327	-,006	-,354	-,250	,595
p <sub>az</sub>	,163	,582	-,118	,120	,580	,884	,477	-,862	-,895	-,093	1,000	,036	-,688	-,232	-,672	,196
p <sub>pl</sub>	,146	-,043	-,070	-,008	-,469	,277	-,672	,065	-,283	-,327	,036	1,000	,079	-,457	-,356	-,845
b <sub>az</sub>	-,261	-,713	,554	-,174	-,506	-,535	-,620	,599	,545	-,006	-,688	,079	1,000	,032	,169	-,242
b <sub>pl</sub>	-,191	-,003	-,016	,013	,379	-,315	,092	-,098	,397	-,354	-,232	-,457	,032	1,000	,625	,048
t <sub>az</sub>	-,186	-,270	-,116	-,041	,034	-,623	-,040	,432	,677	-,250	-,672	-,356	,169	,625	1,000	-,053
t <sub>pl</sub>	-,037	,147	,047	,035	,278	-,101	,674	-,082	,056	,595	,196	-,845	-,242	,048	-,053	1,000

**Results and Discussion:-**

The Cheddok method (Table 3) can be used to describe the significance of the correlation coefficients verbally. For example, there is a close correlation between the NP<sub>1stk</sub> and NP<sub>2stk</sub> parameters, equal to R =-0.820, and there is absent correlation between parameters, MSK and NP<sub>1stk</sub> since the correlation value R = -0.016 is in the range (0; 0.1).

**Table 3:-** Cheddok Correlation Tightness Classification Scale.

	0÷0.1	0.11÷	0.31 ÷	0.51 ÷	0.71÷0.9	0.91	0.991 ÷ 1
--	-------	-------	--------	--------	----------	------	-----------

		0.3	0.5	0.7		÷.99	
Nature of communication	<i>is absent</i>	<i>weak</i>	<i>moderate</i>	<i>noticeable</i>	<i>close</i>	<i>strong</i>	<i>functional</i>

In factor analysis, the correlation matrix is transformed, after which all non-diagonal elements of the correlation matrix turn into zero, and the diagonal elements change their values. This means that the parameters become independent of each other. Most factor analysis methods are based on the principal component analysis (PCA), which converts a set of correlated input parameters into a set of non-correlating factors.

In factor analysis, the correlation matrix is transformed, after which all non-diagonal elements of the correlation matrix turn into zero, and the diagonal elements change their values. This means that the parameters become independent of each other. Most factor analysis methods are based on the PCA, which converts a set of correlated input parameters into a set of non-correlating factors. In this method, after calculating the correlation matrix, its orthogonal transformation is performed and the factor loads are determined from the values of the matrix elements. The most common method of rotation is used - Varimax (non-correlated factors). The method is based on finding the eigenvalues and eigenvectors of the correlation matrix from solving the equation:

$$R = AA' \tag{2}$$

where  $R$  – original correlation matrix;  $A$  - matrix whose elements are factor loadings  $a_{ij}$  parameter  $i$  by component  $j$ ;  $A'$  – transposed matrix. The result of solving this equation is to determine the matrix of factor loadings ( $a_{ij}$ ) (tab.4).

**Table 4:-** Component load matrix.

		Components				Rotated components matrix			
		1	2	3	4	1	2	3	4
1	p <sub>az</sub>	<b>,951</b>	-,138	-,004	-,163	<b>,910</b>	,150	,237	-,208
2	np <sub>2stk</sub>	<b>-,894</b>	,047	,356	,043	<b>-,932</b>	-,093	-,157	-,166
3	np <sub>2dp</sub>	<b>-,859</b>	,436	,087	,184	<b>-,934</b>	,125	-,081	,271
4	np <sub>1dp</sub>	<b>,858</b>	-,427	-,173	-,139	<b>,951</b>	-,156	,075	-,180
5	b <sub>az</sub>	<b>-,795</b>	-,217	-,131	-,289	-,548	-,234	<b>-,651</b>	-,042
6	MW	<b>,709</b>	,222	,274	,438	,392	,178	<b>,796</b>	,002
7	np <sub>1stk</sub>	<b>,706</b>	,376	<b>-,535</b>	,013	<b>,711</b>	,352	,127	,529
8	t <sub>az</sub>	<b>-,541</b>	,486	-,402	,429	-,549	,026	,014	<b>,756</b>
9	p <sub>pl</sub>	-,043	<b>-,918</b>	,159	,308	,072	<b>-,895</b>	,095	-,387
10	t <sub>pl</sub>	,214	<b>,752</b>	,221	-,511	,032	<b>,955</b>	-,034	-,086
11	np <sub>1slip</sub>	,610	<b>,708</b>	,119	-,099	,345	<b>,787</b>	,381	,115
12	np <sub>2slip</sub>	-,095	,403	<b>,775</b>	-,374	-,371	<b>,639</b>	,129	-,591
13	b <sub>pl</sub>	-,161	,513	<b>-,650</b>	,333	-,116	,124	-,013	<b>,891</b>
14	MSK	,257	-,052	<b>,471</b>	,412	-,002	-,095	<b>,626</b>	-,243
15	lnR	-,285	-,250	-,456	<b>-,745</b>	,114	-,009	-,944	-,037
16	azim	,176	,069	,069	<b>,206</b>	,074	,012	<b>,274</b>	,047

Determination of the optimal number of components of the factor model - the number of factors is determined by the number of eigenvalues of the correlation matrix, exceeding one, which are the roots of equation (2). The first colon in table “Cumulative variance explained” (Table 5) shows that four eigenvalues are greater than one, that is, four factors are defined. The attenuation of seismic intensity is an important factor determining the quality of seismic zoning. The calculated seismic intensity values, taking into account the magnitude, epicentral distance and depth, often deviate from the real values of MSK-64 points, at the IDP (Intensity data points) points of the macroseismic field, respectively, the seismic hazard is calculated with errors. In an isotropic geophysical environment, with a uniform distribution of seismic energy, the shape of the theoretical isoseists would be determined by the geometry of the source in the near zone, and the shape close to the circle in the far. However, in reality, from the epicenter of earthquakes, lines of equal intensity diverge in the form of ovals or curved lines of an intricate shape. The shape of

these isolines is influenced by factors: how common is the geometry and mechanism of the source; shear wave velocities, geometric divergence of the wave front; and regional - energy dissipation on inhomogeneities, absorption by the medium, physical properties of the medium, types of soils, features of the geological structure of the medium, composition of the material; the thickness of the soil layer; ground water level; speed differences in bedrock and overlying layers, etc. To take into account the influence of these factors on the intensity of seismic effects, and to improve the quality of seismic hazard maps, more complex models of attenuation of the intensity of their shocks should be made. However, at this stage of development of seismology, the complexity of the attenuation model is limited by the elliptical form of the dependence of the attenuation of the intensity of seismic shocks as the distance from the earthquake source [Rautian 1982; Shebalin, 1961, 1997].

The development of a more complex and adequate attenuation model is the subject of study by many seismologists. For example, an analysis of the macroseismic and instrumental data of earthquakes of the intermediate depth of the Vrancea locus revealed specific features of the effects of earthquakes: the impact on large areas with a predominant orientation of NE-SW; a large degree of dependence of the seismic displacement amplitude of the soil on local and regional geological conditions, as compared with the magnitude and distance from the source; large variability of parameters of strong ground movements; the reflection of the topography of the earth's surface by isolines [Ismail-Zadeh et al., 2007]. In the near zone, where  $r \sim h$ , ( $r$  is the epicentral distance,  $h$  is the depth of the earthquake source), the geometry of the earthquake source has a decisive influence on the configuration of the macroseismic field [Shebalin, 1961; 1980].

**Table 5:-** Cumulative variance explained.

Component	Initial eigenvalues			Extracted sum of squares of loads			Rotated sum of squares of loads		
	Total	% dispersion	Total %	Total	% dispersion	Total %	Total	% Dispersion	total %
MW	<b>5,635</b>	37,567	37,567	5,635	37,567	37,567	5,121	34,138	34,138
azim	<b>3,255</b>	21,697	59,264	3,255	21,697	59,264	3,041	20,272	54,411
np <sub>1stk</sub>	<b>2,157</b>	14,380	73,644	2,157	14,380	73,644	2,354	15,696	70,107
np <sub>1dp</sub>	<b>1,703</b>	11,352	84,996	1,703	11,352	84,996	2,233	14,889	84,996
np <sub>1slip</sub>	,926	6,176	91,172						
np <sub>2stk</sub>	,438	2,920	94,092						
np <sub>2dp</sub>	,298	1,987	96,078						
np <sub>2slip</sub>	,227	1,515	97,593						
P <sub>az</sub>	,180	1,202	98,795						
P <sub>pl</sub>	,079	,525	99,320						
B <sub>az</sub>	,055	,364	99,685						
B <sub>pl</sub>	,031	,204	99,889						
T <sub>az</sub>	,011	,074	99,963						
T <sub>pl</sub>	,004	,030	99,993						
lnR	,001	,007	100,000						

The study of seismicity of Kyrgyzstan revealed that the influence of the source length on the macroseismic field configuration disappears in the interval of 25–35 km, for earthquakes with a magnitude of  $M = 6$ , and for earthquakes with  $M = 7$ , in the interval of 80–110 km. [Januzakov, 2013]. The macroseismic field is characterized by zones of equal intensity and lines separating the points of change of intensity, which are usually quite tortuous, and this fact prompted the addition of a macroseismic field to the equation, characterizing the mechanism and geometry of the earthquake source as seismic variables, to create a more realistic model of seismic intensity attenuation jolts. The apparatus of mathematical statistics has a method of factor analysis; whose task is to reduce the number of regressors in multiparameter problems. Factor analysis is based on determining the minimum number of factors that make up the largest share in the variance of the data. In the study of the complex nature of seismicity, factor analysis helps to better understand the essence of seismic processes, since the interdependence between seismic parameters must be due to relationships between parameters, the identification of which is the task of factor analysis.

Factor analysis allows you to explore the structure of the relationship of variables, where each group of variables will be determined by the factor according to which these variables have maximum loads. The result of factor analysis is a transition from a set of input parameters to a smaller number of variables, called factors. The factor is interpreted as a hidden variable - the cause of the joint variability (interrelation) of several initial parameters. Most factor analysis methods are based on a model of principal components that transforms a group of correlated initial parameters into another group of uncorrelated factors. From the values of descriptive statistics (Table 1) it can be seen that the values of the statistical characteristics of seismic parameters have a large scatter, and their values are initially measured in different units that are incompatible with each other, therefore they should be brought to a single scale using standardization (1). The values of the regression coefficients are estimated using the least squares method, and the Gauss-Markov conditions must be satisfied for the consistency, unbiasedness, and efficiency of the estimates:

$$E(\varepsilon_i) = 0, \quad (i = 1, n) \quad (3)$$

The variance of the random component in all observations should be constant (homoscedasticity) equal to zero, i.e.

$$D(\varepsilon_i) = \sigma^2, \quad (i = 1, n) \quad (4)$$

it is assumed that the random component has a normal distribution i.e.

$$E(\varepsilon_i \varepsilon_j) = 0, \quad \varepsilon_i \sim N(0; \sigma^2), \quad (i \neq j) \quad (5)$$

The multidimensional linear regression model is:

$$y_i = \alpha + \beta x_1 + \beta x_2 + \dots + \beta x_{18} + \varepsilon_i, \quad (i=1, \dots, n) \quad (6)$$

where  $y$  – dependent random variable,  $x_1, \dots, x_j$  – explained (regressors);  $\beta_1, \beta_2, \dots, \beta_j$  – coefficient of regression;  $\varepsilon$  – random variable. The appearance of a random component:

$$\varepsilon_i = y_i - \hat{y}_i \quad (7)$$

where  $y_i$  – real value,  $\hat{y}$  – the values of the dependent variable calculated by the regression equation are related to the effect on the dependent variable  $y$  of unaccounted factors, the inadequacy of the chosen model and errors in the observations.

To explain the 100% variance in seismic data, 16 components would be required. However, using the factor extraction procedure, only 4 factors out of 16 were extracted, which explain 84,996% of the variance in the data set. That is, a factor model consisting of 4 factors saves 84,996% of the initial information. As noted earlier, when grouping the initial array of parameters, the loss of information is inevitable. Saving information by as little as 60% is considered a fairly good indicator. Usually, when conducting factor analysis, the first main components are used, the total proportion of dispersion of which exceeds 60%. Considering that in the course of factor analysis, the number of parameters is reduced several times, even with a large loss of information, 40%, for example, the use of a factor model is appropriate. The initial eigenvalues and eigenvalues after extraction (Extracted sum of squared loads) are the same for PCA (Table 5), however, with other extraction methods they differ slightly. The graph of eigenvalues is useful in determining the number of factors, which is determined by the number of factors where a sharp decline begins. After a given number of factors, the dependence studied is close to the horizontal line, that is, the decrease of eigenvalues slows down. As can be seen (Fig. 1), the inflection point is on the 4th factor, that is, no more than 4 factors stand out. This method is used to determine the number of factors before rotation, the purpose of which is to determine a simple model, where each parameter corresponds to a large value of the factor load for one factor and a small value for all the others. The factor load is determined by the correlation coefficient of each parameter with each of the identified factors. The sum of the squares of the loads on the column represents the variance of the factor and serves as a measure of its information content. The ratio of the variance of the factor to the total variance of the standardized initial variables (numerically equal to their number) makes it possible to judge the relative informativeness of the factor. In the principal component method, the dispersion of the principal components is equal to the eigenvalues of the correlation matrix of the initial variables.

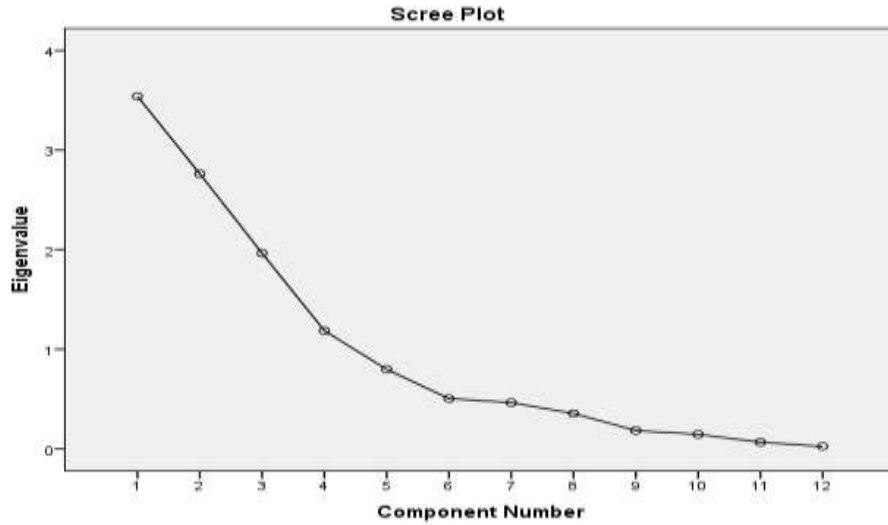


Figure 1:- Graph of the eigenvalues of the correlation matrix.

The component matrix, (table 6) represents the results of factor analysis and the principal component analysis (PCA), the rows of which correspond to the initial variables, and the columns to the factors (main components). At the intersection of the row and the column is the value of the factor load, which is the correlation coefficient between the variable and the factor, which help in interpreting the factors. In PCA, they are used as coefficients for expressing factors through variables in a regression equation. The explanation of the identified factors, and the inclusion of a separate parameter in one factor, is carried out by the magnitude of the correlation of this parameter with this factor (factor load).

Table 6:- Component matrix.

F1	F2	F3	F4
p <sub>az</sub> np <sub>2stk</sub> np <sub>2dp</sub> np <sub>1dp</sub> b <sub>az</sub> M <sub>w</sub> np <sub>1stk</sub> t <sub>az</sub>	p <sub>pl</sub> t <sub>pl</sub> np <sub>1slip</sub>	b <sub>az</sub> MW MSK lnR	t <sub>az</sub> b <sub>pl</sub>

Table 7:- Component Matrix.

		Components				Rotated Component Matrix			
		1	2	3	4	1	2	3	4
1	paz	,951	-,138	-,004	-,163	<b>,910</b>	,150	,237	-,208
2	np <sub>2stk</sub>	-,894	,047	,356	,043	<b>-,932</b>	-,093	-,157	-,166
3	np <sub>2dp</sub>	-,859	,436	,087	,184	<b>-,934</b>	,125	-,081	,271
4	np <sub>1dp</sub>	,858	-,427	-,173	-,139	<b>,951</b>	-,156	,075	-,180
5	b <sub>az</sub>	-,795	-,217	-,131	-,289	<b>-,548</b>	-,234	-,651	-,042
6	MW	,709	,222	,274	,438	<b>,392</b>	,178	,796	,002
7	np <sub>1stk</sub>	,706	,376	-,535	,013	<b>,711</b>	,352	,127	,529
8	t <sub>az</sub>	-,541	,486	-,402	,429	<b>-,549</b>	,026	,014	,756
9	p <sub>pl</sub>	-,043	-,918	,159	,308	,072	<b>-,895</b>	,095	-,387
10	t <sub>pl</sub>	,214	,752	,221	-,511	,032	<b>,955</b>	-,034	-,086
11	np <sub>1slip</sub>	,610	,708	,119	-,099	,345	<b>,787</b>	,381	,115
12	np <sub>2slip</sub>	-,095	,403	,775	-,374	-,371	<b>,639</b>	,129	<b>-,591</b>

13	b <sub>pl</sub>	-,161	,513	-,650	,333	-,116	,124	-,013	<b>,891</b>
14	MSK	,257	-,052	,471	,412	-,002	-,095	<b>,626</b>	-,243
15	lnR	-,285	-,250	-,456	<b>-,745</b>	,114	-,009	<b>-,944</b>	-,037
16	azim	,176	,069	,069	<b>,206</b>	,074	,012	<b>,274</b>	,047

$$X_i = \sum_{j=1}^J a_{ij} F_j, a_{ij} = r_{X_i F_j} \quad (8)$$

Factor analysis is based on the rotation of factors, to obtain a simple structure of the relationship between parameters, where a large value of the factor load of each parameter corresponds to only one factor and a small one for the rest. The “Varimax” rotation option is usually used, which is orthogonal, since with such rotation the axes are located at a right angle. With orthogonal rotation, each successive factor is determined so as to maximize the variance left over from the previous factors, so the factors are uncorrelated with each other. In each row of the rotated matrix of components (Table 7), the factor load is noted, which has the maximum absolute value. For example, the np<sub>1dp</sub> parameter correlates as closely as possible with 1 factor — the correlation value is 0.951.

Parameters; paz; np<sub>2stk</sub>; np<sub>2dp</sub>; np<sub>1dp</sub> are also associated with the first factor of the highest correlation (0,910; -0,932; -0,934; 0,951). In the same way, factors are determined to which the remaining parameters should be attributed (Table 7). When selecting the name of each factor should be based on the logic and research topic. Thus, the parameters: np<sub>1slip</sub>; np<sub>2slip</sub> combined in the first factor (Table 7), hidden from direct measurement, and which causes the interdependence of parameters. Accordingly, the parameters are combined, combined into the remaining four factors. In this case, it becomes possible, instead of the initial set of 16 parameters, to analyze data on the four identified factors. In PCA, the value of each parameter X<sub>i</sub> is represented as linear combinations of factor loads and factors F<sub>j</sub>, j = 1,2, ..., J, J is the number of factors, I is the number of parameters J << I:

The matrix consisting of factor loads and having the number of columns equal to the number common factors, and the number of rows equal to the number of input parameters, is called the factor matrix. The factor matrix captures the degree of linear relationship of each parameter to each common factor. The value of the factor load, close to zero, means that the factor has almost no effect on this parameter. It is possible to calculate the contributions of factors to the total variance of all parameters (by calculating the sum of the squares of the factor loads for each factor for all parameters. The higher the proportion of this contribution in the total variance, the more significant this factor is. Factor estimates are estimates of each parameter for each factor, and indicate emissions by factor (Table 8).

**Table 8:-** Factor scores community.

Parameter	initial	extraction
M <sub>w</sub>	1	0,805
lnR	1	0,858
azim	1	0,139
np <sub>1stk</sub>	1	0,932
np <sub>1dp</sub>	1	0,965
np <sub>1slip</sub>	1	0,897
np <sub>2stk</sub>	1	0,931
np <sub>2dp</sub>	1	0,967
np <sub>2slip</sub>	1	0,919
p <sub>az</sub>	1	0,951
p <sub>pl</sub>	1	0,976
b <sub>az</sub>	1	0,789
b <sub>pl</sub>	1	0,827
t <sub>az</sub>	1	0,871
t <sub>pl</sub>	1	0,923

The matrix of coefficients of factor estimates (Table 8) contains the regression coefficients  $b_{ij}$ , which are used to calculate factor estimates:

$$F_{jn} = \sum_{i=1}^N b_{ij} X_{in}^S \quad (9)$$

where  $F_{jn}$  – value of  $j$ -th factor by  $n$ -th observation,  $X_{in}^S$  – standardized value of  $i$ -th parameter by  $n$ -th observation in initial sample;  $b_{ij}$  – factor loads. In the SPSS package, the calculated factor scores are indicated by identifiers:

FAC1\_1, FAC2\_1, are added to the right of the source data array, and can be used as variables in a further study. In the calculated factor estimates, emissions are not seen in the values of the factors, which indicates a good quality of factor analysis. The sum of the squares of the loads of the  $j$ -th factor in all  $I$  - parameters is equal to the eigenvalue:

$$\sum_{i=1}^I a_{ij}^2 = \lambda_j \quad (10)$$

of this factor, and it is used to characterize the structure of factors, which can be expressed as a percentage, and it helps to know the nature of the factors. One of the characteristics of the quality of the factor model is the commonality - the sum of square factor loadings by the factor matrix row:

$$h_i^2 = \sum_{k=1}^M a_{ik}^2, \quad (11):$$

a given parameter, explained by all factors, and can be interpreted as the reliability of this parameter. The community determines for which parameters factor analysis works better or worse. The magnitude of the community must be greater than 0.5, otherwise the corresponding parameters will be deleted [Inheritances A. 2013]. That is, the community indicates the reliability of this parameter. A value of 1 means that the variance of a parameter is completely determined by the factor to be emitted. Extraction - the proportion of dispersion, which takes values from the interval (0-1), is explained by all the parameters remaining after the extraction of factors. For example, the extracted factors account for over 97.6% of the dispersion of the  $P_{pl}$  parameter (Table 8). If the parameter has low generality, for example, less than 0.5, for this parameter the application of the factor model does not make sense and this parameter should be removed from the model.

When factor analysis under one factor, the parameters of the initial array are collected, which are associated by the closest correlation connection (factorial loads) with this factor, which are contained in the rotated matrix of tab components. Factor loads vary from  $-1$  to  $+1$  and are analogous to the correlation coefficient. In the matrix of factor loadings, it is necessary to distinguish significant loads using student's criterion:

$$t = r\sqrt{N-2} / \sqrt{1-r^2} \quad (12)$$

**Таблица 9:-** Matrix of Reproduced Correlations.

	Stk	Dp	Slip	Stk1	Dp1	Slip1	Az	Pl	Az1	Pl1	Az2	Pl2
Stk	0,786 <sup>a</sup>	0,137	0,207	-0,751	-0,029	-0,054	0,482	-0,023	-0,547	0,259	0,031	-0,127
Dp	0,137	0,716 <sup>a</sup>	-0,215	-0,208	-0,666	-0,205	0,531	0,457	-0,088	-0,193	-0,329	-0,231
Slip	0,207	-0,215	0,739 <sup>a</sup>	-0,283	-0,178	0,661	0,115	-0,715	-0,322	-0,385	0,147	0,757
Stk1	-0,751	-0,208	-0,283	0,758 <sup>a</sup>	0,168	-0,071	-0,549	0,089	0,496	-0,140	0,089	0,008
Dp1	-0,029	-0,666	-0,178	0,168	0,927 <sup>a</sup>	-0,195	-0,536	-0,083	0,202	0,615	0,209	-0,290
Slip1	-0,054	-0,205	0,661	-0,071	-0,195	0,758 <sup>a</sup>	0,022	-0,724	0,029	-0,435	-0,133	0,793
Az	0,482	0,531	0,115	-0,549	-0,536	0,022	0,632 <sup>a</sup>	0,132	-0,316	-0,109	-0,277	-0,033
Pl	-0,023	0,457	-0,715	0,089	-0,083	-0,724	0,132	0,825 <sup>a</sup>	0,091	0,291	-0,113	-0,785
Az1	-0,547	-0,088	-0,322	0,496	0,202	0,029	-0,316	0,091	0,787 <sup>a</sup>	0,109	-0,531	-0,128
Pl1	0,259	-0,193	-0,385	-0,140	0,615	-0,435	-0,109	0,291	0,109	0,706 <sup>a</sup>	-0,079	-0,619
Az2	0,031	-0,329	0,147	0,089	0,209	-0,133	-0,277	-0,113	-0,531	-0,079	0,874 <sup>a</sup>	0,123

PI2	-0,127	-0,231	0,757	0,008	-0,290	0,793	-0,033	-0,785	-0,128	-0,619	0,123	0,942 <sup>a</sup>
Matrix residual reproduced correlations												
Stk		-0,043	0,007	0,025	-0,040	-0,078	-0,081	0,031	0,144	-0,111	0,021	0,033
Dp	-0,043		0,058	0,076	-0,021	0,076	-0,073	-0,148	-0,015	0,230	0,083	-0,043
Slip	0,007	0,058		0,097	-0,010	-0,100	-0,026	0,022	0,013	0,057	-0,030	-0,054
Stk1	0,025	0,076	0,097		0,027	-0,016	0,163	-0,082	-0,029	0,123	0,016	-0,024
Dp1	-0,040	-0,021	-0,010	0,027		0,015	0,083	0,021	-0,056	-0,036	-0,031	0,001
Slip1	-0,078	0,076	-0,100	-0,016	0,015		-0,003	0,012	-0,079	0,087	0,052	-0,060
Az	-0,081	-0,073	-0,026	0,163	0,083	-0,003		-0,058	-0,022	0,049	0,042	0,007
Pl	0,031	-0,148	0,022	-0,082	0,021	0,012	-0,058		-0,011	-0,160	-0,071	-0,006
Az1	0,144	-0,015	0,013	-0,029	-0,056	-0,079	-0,022	-0,011		-0,052	0,104	0,026
Pl1	-0,111	0,230	0,057	0,123	-0,036	0,087	0,049	-0,160	-0,052		0,086	-0,046
Az2	0,021	0,083	-0,030	0,016	-0,031	0,052	0,042	-0,071	0,104	0,086		-0,010
PI2	0,033	-0,043	-0,054	-0,024	0,001	-0,060	0,007	-0,006	0,026	-0,046	-0,010	

**The criterion for the correct choice of a number of factors:**

The matrix of differences between the reproduced and real correlations (tab. 9) is the matrix of residues of the reproduced correlations, which is used to identify certain correlations that are poorly reproduced by the model with a given number of factors. Insignificant residual values indicate the correctness of the choice of a number of factors. This is the correlation matrix of the original items, which would have turned out if we assume that the calculated factors are correct and the only factors. For any pair of variables taken, the reproduced correlation is equal to the product of their factor loads on the first factor plus the product of such loads on the second factor, etc. for all factors. Remains of reproduced correlations or “matrix of correlation residues” is a matrix of differences between reproduced and real correlations. The insignificant remains, the more likely the decision in the choice of four factors. The matrix of residuals of the reproduced correlations helps to identify specific correlations that are poorly reproduced by the factor model with a given number of factors.

KMO criterion is used to check the quality of data factorization based on correlation and partial correlation between the initial parameters. Sample statistics KMO is also used to identify multi collinearity in factor analysis to determinewhich parameters need to be removed from the model. In this case, the sample KMO index value of 0.57 (tab. 10) exceeds the threshold value of 0.5, which indicates the suitability of the data for conducting factor analysis. Testing hypothesis H<sub>0</sub>: that the parameters involved in factor analysis are not correlated with each other, with the alternative hypothesis of the presence of correlation, is carried out using the Bartlett criterion. The probability of the significance of the sample statistics of the Bartlett test criterion (significance line) equal to 0.0, below the significance level  $\alpha = 0.05$ , therefore, the hypothesis H<sub>0</sub>, is rejected in favor of an alternative hypothesis about the correlation of the parameters among themselves. Thus, the test criteria of KMO and Bartlett confirm the correctness of the decision on the use of factor analysis to study the totality of the values of seismic parameters.

**Table 10:-** KMO and the Bartlet sphericity criterion.

A measure of the selective adequacy of the Kaiser-Meier-Olkin (KMO)		0,571
Bartlett Sphericity Criterion	Custom Chi-square statistic	776,147
	degrees of freedom	66
	significance	,000

For factor analysis, the total KMO measure must be greater than 0.50. If KMO is less than the threshold value, then it is necessary to remove the parameters with the lowest values of the statistics of the individual KMO. The process should be repeated until the total KMO, which is the sum of the KMO statistics for each parameter, reaches the threshold value. KMO private statistics for each parameter are diagonal elements of the Anti-image matrix (Table 11).

**Table 11:-** Anti-image matrix.

	Stk	Dp	Slip	Stk1	Dp1	Slip1	Az	Pl	Az1	Pl1	Az2	Pl2
Stk	0,331	0,004	-0,103	0,224	-0,006	0,095	-0,117	0,012	-0,050	0,003	-0,076	0,011
Dp	0,004	0,178	0,023	-0,039	0,127	-0,017	0,001	0,025	0,054	-0,029	0,053	0,029
Slip	-0,103	0,023	0,342	-0,063	0,027	-0,014	0,056	0,035	0,121	0,000	0,084	-0,010
Stk1	0,224	-0,039	-0,063	0,320	-0,038	0,078	-0,052	0,007	-0,152	0,022	-0,133	0,003
Dp1	-0,006	0,127	0,027	-0,038	0,123	-0,006	0,057	0,022	0,044	-0,037	0,037	0,020
Slip1	0,095	-0,017	-0,014	0,078	-0,006	0,335	-0,003	0,035	0,013	0,006	0,079	-0,014
Az	-0,117	0,001	0,056	-0,052	0,057	-0,003	0,516	0,020	0,166	-0,003	0,180	0,009
Pl	0,012	0,025	0,035	0,007	0,022	0,035	0,020	0,075	0,033	0,066	0,046	0,048
Az1	-0,050	0,054	0,121	-0,152	0,044	0,013	0,166	0,033	0,399	0,003	0,278	0,013
Pl1	0,003	-0,029	0,000	0,022	-0,037	0,006	-0,003	0,066	0,003	0,118	0,013	0,053
Az2	-0,076	0,053	0,084	-0,133	0,037	0,079	0,180	0,046	0,278	0,013	0,480	0,014
Pl2	0,011	0,029	-0,010	0,003	0,020	-0,014	0,009	0,048	0,013	0,053	0,014	0,044

Multicollinearity means high mutual correlation of explanatory regression variables. The absence of high collinearity of regressors is one of the conditions for applying the least squares method for estimating the parameters of multidimensional linear regression. The study of seismicity is associated with the identification of cause-effect relationships in endogenous processes. In the study of the relationships between causal investigative processes and the determination of the strength of the corresponding relationships, regression analysis may be useful, which makes it possible to identify trends in mass phenomena. Regression analysis assumes that the following conditions are met:

The relationship between the variables is linear, which can be seen on scatter diagram of variable values.

Normal distribution of residuals - the difference between the predicted and observed values. Visually determine the nature of the distribution of residues can be histogram of residues, and its comparison with the density of the standard normal distribution. It is believed that two regressors  $x$  and  $z$  are collinear if the pair correlation coefficient  $r_{xz} > 0.7$  [Gabrielyan, 2006]. The absence of multicollinearity and homoscedasticity - consistency of the variance of the regression residuals is a necessary condition for obtaining unbiased, effective and consistent estimates of the regression coefficients. Checking the model for collinearity of regressors is determined by the values of the criterion:

$$VIF_j = \frac{1}{1 - R_j^2}$$

$$R_j^2 = \frac{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2}{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2} \quad (13)$$

Where  $R_j^2$  - multiple determination coefficient,  $n$  - number of observations;  $x_{ji}$  - observed values,  $\bar{x}_j$  - calculated values and the arithmetic average of the dependent variable is in the regression of the variable  $X_j$  from to other variables.

$$x_j = c_0 + c_1 x_1 + \dots + c_{j-1} x_{j-1} + c_{j+1} x_{j+1} + \dots + c_k x_k + \varepsilon, j = 1, \dots, m \quad (15)$$

In practical applications, it is recommended that the number of observations  $n$  be greater than the number of independent variables  $m$  not less than three times [Mkhitarayan, et al., 2008]. High values of  $VIF_j$  - greater than 4, at least for one regressor  $X$ , means multicollinearity.

The multicollinearity value will be indicated by the values of the VIF coefficient (13) from 4 and higher for at least one  $j$ . The quality characteristic of the regression model is the coefficient of determination of the R-square, which is



equation, and using test criteria, the adequacy of the regression is checked. If the model is significant, the next regressor is included, and so on, if the significance of some regressor is less than the critical value, then the regressor is excluded, if more, then those associated with the dependent variable with the largest partial correlation are preserved and are included in the regression equation. In this case, direct and inverse step-by-step regressions are combined, where the parameters are successively turned on and off in the equation. The multiple regression method allows determining the significance of a linear relationship between parameters and seismic intensity, the quality of data approximation by a regression equation, the suitability of the calculated values of the coefficients of the equation for the best prediction, as well as determining the significance of parameters for predicting the seismic force. From the correlation matrix of all sample data it can be seen that a noticeable correlation is observed between the intensity of seismic shocks and the hypocentral distance; moderate dependence between magnitude and MSK-64 scale points. A functional and strong relationship was observed between some of the physical parameters of the foci, probably due to endogenous processes, the nature of which is to be clarified. The table shows that the method of step-by-step regression in the regression equation alternately included the parameters: Constant, lnR MW, np2slip, tpl, azim, starting with a constant and hypocentral distance. In the pivot table (tab. 8), the values of the coefficient of determination for each model, which takes the highest value  $R^2 = 0.823$ , are given for the model, where the regressors are included in the equation: Constant, lnR  $M_w$ , np2slip, tpl, azim, this means that the regression accounts for about 82.3% of the percent seismic intensity variation. The Durbin-Watson coefficient checks the autocorrelation, takes values from the interval (0.4). Values close to 0 indicate strong positive autocorrelation, close to 4, strong negative autocorrelation, and close to 2 indicate no autocorrelation. In practice, if the condition from  $1.5 < d < 2.5$  is met for the value of the Durbin-Watson coefficient, there is no autocorrelation. The observed value is 1.158, therefore, we can talk about the presence of positive autocorrelation.

**Table 12:-** Model summary table.

Model	R	Coefficient determinations R-squared	Adjusted R-squared	Standard mistake ratings	Change Statistics					Durbin-Watson
					square change	F Change	df1	df2	sig. F Change	
1	,836 <sup>a</sup>	,699	0,699	0,715	0,699	2421,724	1	1043	,000	
2	,886 <sup>b</sup>	,784	0,784	0,606	,0085	411,105	1	1042	,000	
3	,897 <sup>c</sup>	,804	0,804	0,577	,020	106,931	1	1041	,000	
4	,903 <sup>d</sup>	,816	0,815	0,560	,012	66,228	1	1040	,000	
5	,907 <sup>e</sup>	,823	0,822	0,549	,007	41,741	1	1039	,000	1,158
a. Predictors: (Constant), lnhras										
b. Predictors: (Constant), lnR MW										
c. Predictors: (Constant), lnR MW, np2slip										
d. Predictors: (Constant), lnR MW, np2slip, tpl										
e. Predictors: (Constant), lnR MW, np2slip, tpl, azim										
f. Dependent Variable: I										

According to Fisher's criterion, the regression model turned out to be significant for all included in 5 regressors models, since the probability of the significance of the F – criterion sampling statistics used to test the null hypothesis  $H_0$ : that the regression equation is insignificant is zero. The best, according to statistical characteristics, was the model, which included 5 parameters: constant,  $M_w$ , lnR np2slip, tpl and azim. In this case, the coefficient of determination is greatest:  $R^2 = 0.823$ . The sample distribution function of the regression residuals (Fig. 2) almost coincides with the theoretical distribution function, which speaks in favor of the model with 5 regressors. However, test criteria characterizing the adequacy of the regression model indicate sufficient statistical significance of the attenuation model, where the constant, magnitude  $M_w$ , and the natural logarithm of the hypocentral distance lnR are included. The inclusion of other parameters in the model did not lead to a significant improvement in the quality of the attenuation model, and such parameters are difficult to statistically process. It was decided to include in the regression equation the attenuation of magnitude and distance, as is customary in seismology. In this task, the values of the parameters are given in different units of measurement, therefore, to estimate the degree of influence of parameters on the intensity of seismic shocks, standardized coefficients in the regression equation are used, that is, a standardized version of the regression equation. Instead of the values of all parameters, their standardized values are used. In this case, only the form of the regression equation record changes, the correlation coefficients between all

parameters, the coefficient of determination, the significance of the regression coefficients do not change. The probability of the significance of the t-statistic is checked null hypothesis about the insignificance of the regression coefficient. The table shows that in the third, fourth and fifth models, the coefficient “Constant” is statistically insignificant - the probability of significance exceeds 5%, the significance level  $\alpha = 0.05$ . Values of the VIF coefficient less than 4, therefore, multicollinearity is absent. One of the indicators of collinearity of parameters is the tolerance (tolerance) of a parameter, the sample values of which significantly exceed the threshold value of 0.2. Therefore, the corresponding parameter is not a linear combination of other regressors. One of the prerequisites determining the adequacy of the regression model is the normal distribution of residuals, which is indicated by those who slightly deviate from the distribution density and the distribution function of the normal law (Fig. 2, a; b), the histogram and the sample residual distribution function.

The approximation by an ellipse of a macroseismic field is a simplified model of the real picture, which is mosaic, and the lines dividing zones of equal magnitude - meandering. To build a more realistic model of attenuation of the intensity of seismic shocks, as the distance from the earthquake source increases, a scheme is proposed: the origin of coordinates is combined with the epicenter of the earthquakes under study; in each sliding azimuth section of a width of  $40^\circ$ , with a step of  $20^\circ$ , the values of the coefficients of the macroseismic field equation are determined using regression analysis (table 15). Built a regression model of the intensity of seismic effects in points-points: on the magnitude of  $M_w$  and log hypocentral distance  $\ln R$  and all parameters  $M_w$ ,  $\ln R$ ,  $azim$ ,  $np_{1stk}$ ,  $np_{1dp}$ ,  $np_{1slip}$ ,  $np_{2stk}$ ,  $np_{2dp}$ ,  $np_{2slip}$ ,  $p_{az}$ ,  $p_{pl}$ ,  $b_{az}$ ,  $b_{pl}$ ,  $t_{az}$ ,  $t_{pl}$ . Test criteria for verifying the significance of a regression indicate a sufficiently high-quality regression of the values of the intensity of seismic jolts by 2 parameters:  $M_w$ ,  $\ln R$ . The regression of intensity values on magnitude and hypocentral distance is characterized by selective values:  $R^2 = 0.784$ , a high value of Fisher's test criterion, which rejects the hypothesis that regression is not significant in favor of its significance.

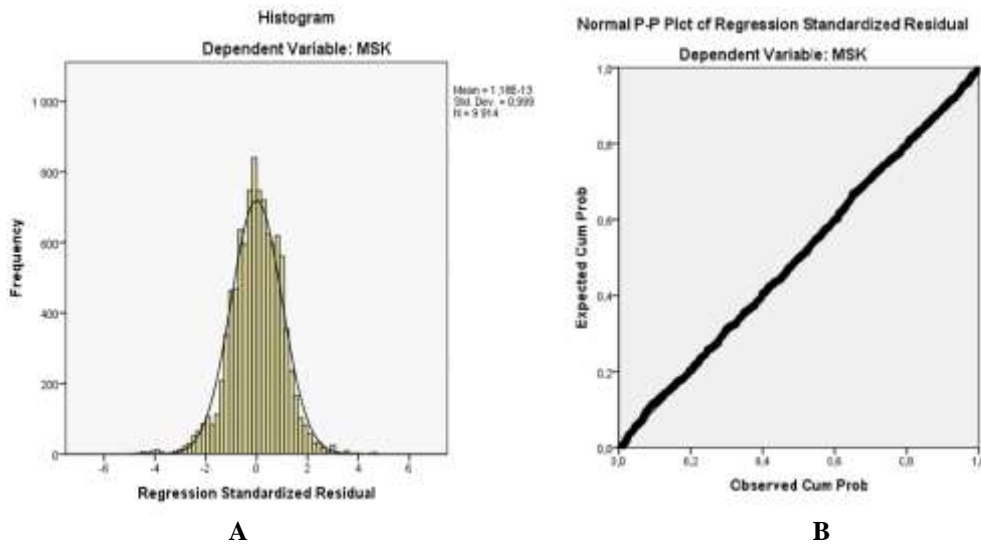


Figure 2:- a) histogram of regression residuals and a plot of the density of the standard normal distribution; b) Theoretical and selective distribution function of residual regression.

A sliding azimuth range of width  $40^\circ$ , in increments of  $20^\circ$ , is used to determine the value of the coefficients of the macroseismic fields equation (table.13)

Table 13:-

Stvor	Regressors			Statistical characteristics			
	Constant	Mw b	Inhras c	Coefficient of determination	Probability of significance F-statistics	Index value VIF	Index Durbin- Watson
0-40 <sup>0</sup>	9.56	0.96	4.40	0.784	0.0	1.011	1.007
20 <sup>0</sup> -60 <sup>0</sup>	8.84	1.06	-4.36	0.805	0.0	1.003	1.014
40 <sup>0</sup> -80 <sup>0</sup>	9.01	1.05	-4.42	0.798	0.0	1.026	0.719

60 <sup>0</sup> -100 <sup>0</sup>	11.20	0.94	-5.10	0.767	0.0	1.005	0.691
80 <sup>0</sup> -120 <sup>0</sup>	10.27	1.13	-5.30	0.743	0.0	1.002	0.346
100 <sup>0</sup> -140 <sup>0</sup>	6.28	1.29	-4.01	0.430	0.0	1.37	0.379
120 <sup>0</sup> -160 <sup>0</sup>	6.03	1.37	-4.05	0.501	0.0	1.394	0.314
140 <sup>0</sup> -180 <sup>0</sup>	5.91	1.53	-4.39	0.774	0.0	1.748	0.403
160 <sup>0</sup> -200 <sup>0</sup>	7.19	1.48	-4.73	0.476	0.0	1.215	0.608
180 <sup>0</sup> -220 <sup>0</sup>	9.33	1.27	-4.94	0.533	0.0	1.228	0.652
200 <sup>0</sup> -240 <sup>0</sup>	9.33	1.27	-5.02	0.546	0.0	1.159	0.666
220 <sup>0</sup> -260 <sup>0</sup>	4.94	1.64	-4.60	0.495	0.0	1.124	0.659
240 <sup>0</sup> -280 <sup>0</sup>	3.30	2.04	-5.40	0.466	0.0	1.311	0.726
260 <sup>0</sup> -300 <sup>0</sup>	7.10	1.61	-5.88	0.618	0.0	1.321	0.897
280 <sup>0</sup> -320 <sup>0</sup>	9.78	1.05	-5.16	0.612	0.0	1.162	0.595
300 <sup>0</sup> -340 <sup>0</sup>	9.17	1.04	-4.72	0.605	0.0	1.016	0.596
320 <sup>0</sup> -360 <sup>0</sup>	7.86	1.091	1.977	0.566	0.0	1.02	0.710
340 <sup>0</sup> -20 <sup>0</sup>	9.14	0.914	1.742	0.697	0.0	1.022	1.14

The observed value of Index Durbin-Watson is 1.014 (table,13), therefore, we can talk about the presence of weak positive autocorrelation. Values of the VIF coefficient are less than 4, therefore there is no multicollinearity. A seismic hazard was calculated and a map of seismic zoning of the territory of Moldova and neighboring countries was constructed based on the obtained values of the attenuation function coefficients of (fig.3, 4). The methodology for calculating seismic hazard developed at the Institute of Geology and Seismology was applied. The mapped territory was covered by a geographical network with a discretization step of 0.2<sup>0</sup>, and in the network nodes the seismic hazard was calculated using the formula [Burtiev, 2017]:

$$p_{\varphi,\psi}(t, n, m_k, I^k) = p(t, n) \frac{n!}{m_k!(n - m_k)!} \omega_k^{m_k} (1 - \omega_k)^{n - m_k} \quad (19)$$

where  $\omega_k$  - the probability of occurrence in the nodes of the grid seismic shock with intensity  $I^k$ , which is the vector of the probability distribution of occurrence in the nodes of the network of seismic tremors with intensity  $I^k=1,2,\dots,12$  degree of the scale MCK-64:

$$\vec{\omega} = (\omega_1, \omega_2, \dots, \omega_{12}) \quad (20)$$

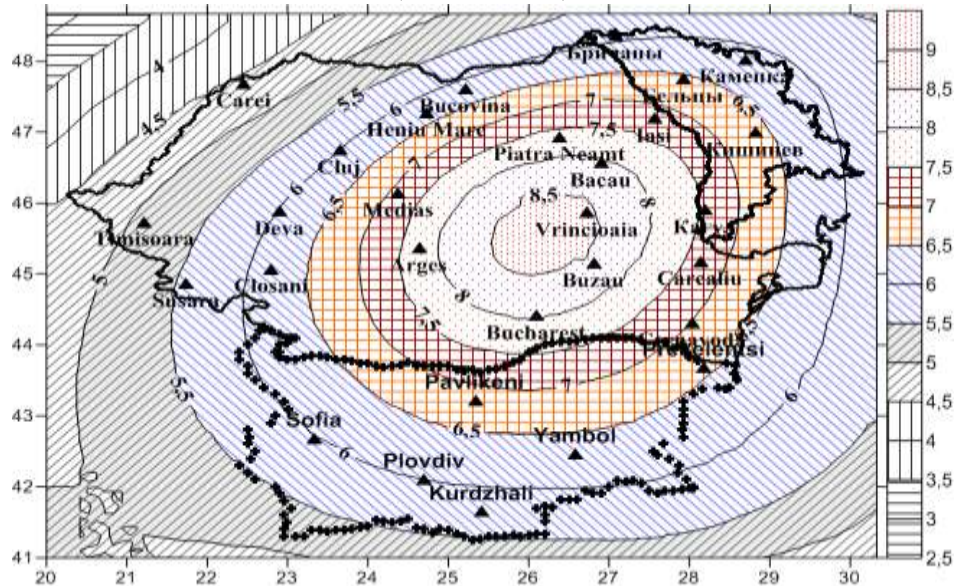


Figure 3:- Seismic zoning map of the territory earthquakes of the Vrancea focus.

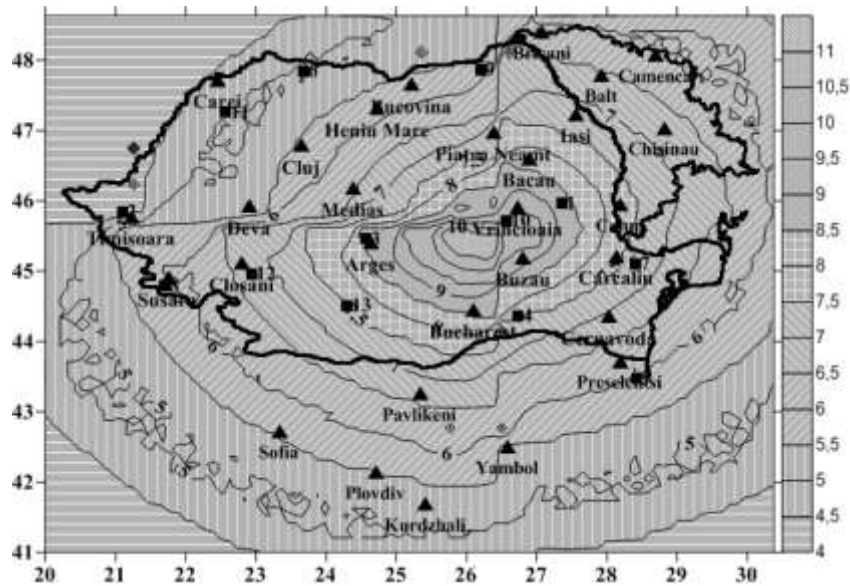


Figure 4:- Seismic zoning map of the area exposed to Romanian earthquakes.

### Conclusion:-

In the study of the complex nature of seismicity, factor analysis turned out to be useful for understanding the essence of seismic processes. A statistically significant correlation dependence was observed between the 16 parameters, reflecting the mechanism and geometry of the Vrancea source, and using factor analysis, hidden factors were found that determine the relationships between these parameters. Factor analysis leads to a loss of information contained in the source data, but a significant reduction in the number of parameters justifies its use and helps to identify patterns in seismic processes that are not directly observable. The seismic hazard map of the territory of Moldova, Romania and Bulgaria in points of the MSK-64 scale, according to the values of the attenuation function coefficients (Table 6), calculated using the moving average method from all seismic zones.

### Recommendations:-

A method for calculating the values of the attenuation equation coefficients is proposed. On the basis of the obtained values, the seismic hazard is calculated, as the probability that at points of the earth's surface at a fixed time  $t$ ,  $n$  pushes happen, of which  $m$  times with intensity  $I^k$ ,  $k$  degree of MSK 64 scale. In order for the regression analysis based on the usual least squares method to give the best results, the random error must satisfy the Gauss-Markov conditions: the mathematical expectation of the random error in any observation must be zero, which means it should not have a systematic bias. Usually, if the regression equation includes a free term, then this means that the condition is satisfied automatically, since the role of the constant is to determine any systematic tendency of the explained variable included in the regression equation. Multicollinearity means a high cross-correlation of the explanatory regression variables. The lack of high collinearity of the regressors is one of the conditions for applying the least squares method to estimate the parameters of multidimensional linear regression. To assess the values of the coefficients of the attenuation function, in the presence of multicollinearity, we use regression analysis on the main components, where the strongly correlated regressors are replaced by components F1, F2, F3, F4, identified by the model of the main components of factor analysis, between which there is no correlation.

### References:-

1. A.I. Bakhtin, N.M. Nizamutdinov, N.M. Khasanova, E.M. Nureyeva, factor analysis in geology
2. Kazan, 2007, 32.
3. Burtiev Rashid, Seismic Hazard Assessment Method Based on the Stochastic Models of Seismicity, Bulletin of the International Institute of Seismology and Earthquake Engineering,
4. Vol.51, 2017, Tsukuba, Japan, 22-38.
5. R. G. Gabrielyan, Econometrics, Vidnoe, 2008, 85.
6. K. D. Januzakov Regional Attenuation Features intensity of shocks of strong earthquakes of Kyrgyzstan and adjacent areas. Bulletin of the Institute of Seismology, NAS KR No. 2. Bishkek. 2013, 11–16.

7. Ismail-Zadeh, A., Sokolov, V. and Bonjer, K., Tectonic stress, seismicity, and seismic hazard in the southeastern Carpathians, *Nat. Hazards*, Springer, 2007, 1-22.
8. Lindsay I Smith, *A Tutorial on Principal Components Analysis*, 2002, 26.
9. Kronrod Tatiana, Panza Giuliano, Radulian Mirchea: Integrated transnational macroseismic data set for the strongest earthquakes of Vrancea (Romania). [issues.pangaea.de/browse/PDI-3917](http://issues.pangaea.de/browse/PDI-3917),
10. *Tectonophysics*, 2012, 271–282
11. Mkhitarian V.S., Arkhipova M.Yu., Sirotin V.P., *Econometrics*, Moscow, 2008, 142
12. Inheritances A. *IBM SPSS Statistics 20 and AMOS: Professional Statistics data analysis*, St. Petersburg,
13. 2013. Electronic textbook on statistics StatSoft.Main components and factor analysis. URL: [ttp://www.statsoft.ru/home/textbook/modules/stfacan.html](http://www.statsoft.ru/home/textbook/modules/stfacan.html)
14. Harman G. *Modern factor analysis*. M., 1972.
15. Rautian T.G. et al. Macroscopic manifestation of earthquakes and focal spectra. *Engineering seismology issues*; issue 23. Moscow. The Nauka, 1982. 74–82.
16. Shebalin N.V. The intensity, magnitude and depth of the earthquake source. *Earthquakes in the USSR*, Moscow, Science, 1961, 126–138.
17. Shebalin, *Seismic zoning of the USSR*, 1980.