Professor Marian DÂRDALĂ, PhD
The Bucharest University of Economic Studies
Associate Professor Titus Felix FURTUNĂ, PhD
The Bucharest University of Economic Studies
Lecturer Cristian IONIȚĂ, PhD
The Bucharest University of Economic Studies

STATISTICAL METHODS FOR ANALYZING THE RELATIONSHIP AMONG ETHNIC, RELIGIOUS AND LINGUISTIC DIVERSITY IN ROMANIA

Abstract. *This paper employs and adapts statistical methods in order to discover particular and general issues among ethnic, religious and linguistic phenomena at the Romanian level. In this sense it is necessary first to determine the extent to which the three aspects are correlated at a general level. Generalized canonical analysis was used to highlight the simultaneous relations among ethnic, religious and linguistic diversity. This analysis is continued by underlying the particular aspects of the relation at the Local Administrative Units (level 2) – LAU2. In order to reflect these issues there were used the canonical ranks, the ranks correlations and the models variables as well as the differences among the diversities indexes computed for ethnic groups, religious and mother tongues.*

Key words: *canonical analysis, correspondence analysis, diversity indexes, ethnic, religious, mother tongues.*

JEL Classification: C38, Z13

1. Introduction

The issue of ethnic diversity is now tackled in close connection with social cohesion, trust and social participation among people with different backgrounds at local community level. Current studies highlight the adverse and benefic social and economic influences of ethnic diversity. Putnam concludes in [8] that a large ethnic diversity leads to the decreasing of social capital. In [14] it is argued that ethnic diversity usually reduces the trust between people of different ethnicities. Merlin Schaeffer in [9] presents a completed analysis about the effects of ethnic diversity. His book exposes the contradictory effects of statistical diversity on the behavioral and cognitive aspects of social cohesion. It also treats social aspects of the interaction between different ethnic groups. Other papers analyze the relationship between diversity and economic performance and the level of living [1].

This paper doesn't aim to analyze ethnic, linguistic and religious diversity considering aspects of social, cohesion, interaction, mutual trust or economic standard. The main goal is to identify relationship pattern among ethnicity, mother tongue and confession in Romania. Another aspect is to analyze how ethnic communities have preserved their individuality. This is highlighted by the relationship between ethnicity on the one hand and mother tongue or confession on the other hand.

The case study and analysis in this paper are based on data collected at the last Romanian's census from 2011 [19]. R software was used in order to process data and displaying the results. Working with spatial data involved specialized packages usage, more technical details can be found in [3].

2. The generalized canonical analysis of the relationship among ethnic, religious and linguistic diversity

In generalized canonical analysis, the linkage between several sets of data is studied. Generalized canonical analysis highlights what a group of phenomena have in common from an informational point of view. Generalized canonical analysis can be accomplished by building up functions that optimize the linkage between canonical variables of groups. Some of the optimization methods are compared and analyzed in Kettenring[6] and Gower [4]. From the many implementations known in the literature, the RGCCA model proposed by Arthur and Michel Tenenhaus[16] is used in this study. According to this model, successive sets of canonical variables are calculated, one for each variables group by solving the following optimization problem:

$$\underset{a_1,\ldots,a_m}{Maximize} \sum_{j,k=1,\, j\neq k}^{m} c_{jk}\, g\left(Cov\left(X_j a_j^s, X_k a_k^s\right)\right),$$

subject to the constraints: $\left\|a_j^s\right\|^2 = 1, \quad j=\overline{1,m}, \left(a_j^l\right)^t a_j^s = 0, \quad l=\overline{1,s-1}, j=\overline{1,m}$

where m is the number of groups, X_j, $j=\overline{1,m}$, the matrices that represent the groups of variables observed on the same set of individuals, c_{jk} $j=\overline{1,m}$, $k=\overline{1,m}$, the functional relationships between variable groups, a_j^s, $j=\overline{1,m}$, the weight vectors for each group and s root, $z_j^s = X_j a_j^s$, and g is the sign function (identity, square and absolute value).

For the three aspects, ethnic, religious and linguistic, three sets of data at the LAU2 level are set up concerning number of people based on ethnicity, religion and mother tongue: X_1, X_2, X_3. Appling the generalized canonical analysis of the three sets of data, great correlations between the canonical variables of the three groups were obtained (Table 1).

Table 1. The lists of the first five canonical roots

Canonical root	$R(z_1, z_2)^2$	$R(z_1, z_3)^2$	$R(z_2, z_3)^2$
Root 1	0.886518182	0.994103438	0.877564571
Root 2	0.655166262	0.974313322	0.645515914
Root 3	0.827764215	0.977838289	0.828511582
Root 4	0.669259562	0.924314419	0.603081002
Root 5	0.808737126	0.946824623	0.807373271

The number of significant canonical roots can be determined by a Bartlett-Wilks test similar to that used in the regular canonical analysis[12]. For the 14 canonical roots, the *pValues* significance thresholds are less than 0.01 (Table 2). This indicates a strong link between the three phenomena, ethnic, religious and linguistic.

Table 2. **Bartlett-Wilks test of significance**

No of root	Chi2	Degree of freedom	PValues
1	14573.29945	21	0
2	9868.846537	20	0
...			
13	32.93963508	9	0.000136823
14	31.50474239	8	0.000114212
...			
21	0.018426055	1	0.892024765

Figure 1 presents the diagram of the structural relation among the three phenomena. This model is applied considering the idea that the relation between linguistic and religious elements is not a direct one.

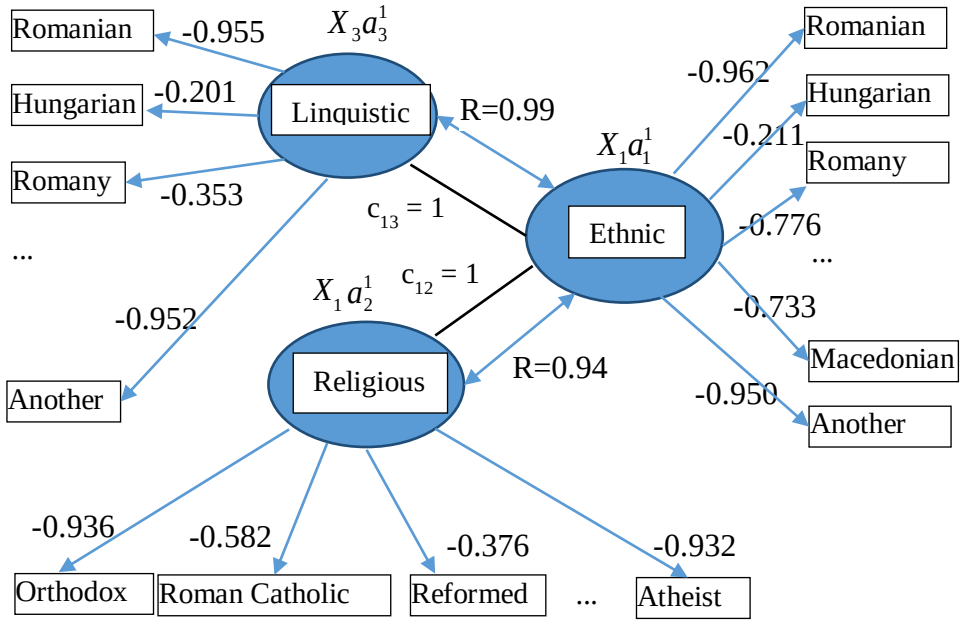

Figure 1. RGCCA applied - horst scheme

At the level of the whole country, the existence of a strong link between ethnic, religious and linguistic phenomena is evident and it was expected to be confirmed by statistical methods. If this link is analyzed at the level of administrative units, there can be found regions or communities which don't confirm the relationship that exists at the country level. The purpose of the analysis is to discover units with particular associations that don't confirm the general pattern of the relationship between the ethnic and linguistic characteristics or between the ethnic and religious characteristics.

The analysis of the discrepancies according to the general pattern can be done by calculating the differences between canonical scores at unit level between set X_1 (Ethnic) and X_2 (Religious), respectively between set X_1 and X_3 (Linguistic), thus:

$$d_i = \sqrt{\prod_{s=1}^{k}\left(z_{1i}^s - z_{2i}^s\right)^2}\ ,\ i=\overline{1,n}\ ,$$

where n is number of units, z_{ji}^s are the canonical variates for set X_j, unit i, root s and k is the number of significant canonical roots.

For greater accuracy, these differences can be calculated for canonical scores obtained by applying separate canonical analysis for sets X_1 and X_2, respectively X_1 and X_3.

Figure 2 shows the maps with the distribution of the scores differences at the Romanian level.

Figure 2. Canonical scores differences

3. The analysis of correspondences in the ethnic-religious and ethnic-linguistic contingency tables

The results of the last census in Romania (2011) are synthetically presented in the form of contingency tables at national level and by residential (urban and rural), where the cross-sectional frequencies for the ethnic and religious variables, respectively ethnic and native language, are stored. These data allow us to perform a detailed analysis of the relationships between the ethnic element and the religious and linguistic elements on the other hand, using the correspondence analysis. Two analyses will be performed, one to study the relationship between ethnicity and religion, and another one to study the relationship between ethnicity and mother tongue. In both analyzes the rows represent the ethnic groups and the columns represent the religions for the first analysis and the mother tongues for the second one.

Correspondence analysis is closely linked to the Chi-square independence test. The Chi-square test is used to determine whether the variables are independent of one another, or in other words, whether there is a statistically significant dependence between them. The Chi-square statistic is calculated as follows:

$$\chi^2 = n \cdot \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{\left(f_{ij} - f_{i.} f_{.j}\right)^2}{f_{i.} f_{.j}},$$

where n is the number of subjects (total population), f_{ij} are the relative frequencies, $f_{i.}$, $f_{.j}$ are the marginal frequencies, p is the number of values for first variable (number of ethnicities) and q is the number of values for second variable (number of religions or languages).
The results of the test for the relationship between ethnicity-religion and ethnicity-mother tongue are presented in Table 3:

Table 3. Chi-square test of independence between ethnicity and religion/language

	Religion	**Language**
Chi-square	58711667	279175533.7
Degree of freedom	420	420
P-value	0	0

The results presented in table 3 denote a strong relationship between variables in the both cases. In the both cases *P-value* is almost 0. A higher Chi-square value means a stronger link, thus, the ethnicity-language relationship is stronger, because of the chi-square value is higher.

The amount of $I = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\left(f_{ij} - f_{i.}f_{.j}\right)^2}{f_{i.}f_{.j}}$ represent the inertia of data.
Corresponding analysis decomposes the total inertia of data on orthogonal axes, each axis taking over some of the inertia. The tables 4 and 5 show the distribution of inertia on the axes, with the emphasis of the axes which cover over 90% from the total inertia, for each analysis.

Table 4. Axes inertia for ethnicity-religion relationship

Axes	**Inertia**	**Cumulative**	**Percent**	**Cumulative**
1	0.76295	0.76295	24.49818	24.49818
2	0.725345	1.488293	23.29078	47.78897
...				
5	0.39012	2.86242	12.52677	91.91203
...				
20	1.5E-09	***3.114302***	4.82E-08	100

Table 5. Axes inertia for ethnicity-language relationship

Axes	Inertia	Cumulative	Percent	Cumulative
1	0.991103	0.991103	6.703945	6.703945
2	0.941511	1.932614	6.368496	13.07244
...				
17	0.507298	13.74181	3.431425	92.95132
18	0.499001	14.24082	3.375305	96.32662
19	0.383453	14.62427	2.593723	98.92034
20	0.159615	***14.78388***	1.079656	100

The detailed analysis of the link between the variables is done by determining the contribution of each frequency to the total inertia or to the total value (χ^2). A frequency reflects the link between two modalities, one for each variable. The contribution of a frequency to the total inertia is $c_{ij} = \frac{\left(f_{ij} - f_{i.} f_{.j}\right)^2}{f_{i.} f_{.j}}$ and contribution to χ^2 is $c_{ij} \cdot \sqrt{n}$. These values represent the square of the deviations of each frequency from the assumption of total independence between the variables. Deviations from the hypothesis of independence can be calculated as follows:

- $r_{ii} = \frac{\left(f_{ij} - f_{i.} f_{.j}\right)}{\sqrt{f_{i.} f_{.j}}}, \quad i = \overline{1, p}, \quad j = \overline{1, q}$ as Pearson residuals,
- $s_{ii} = r_{ij} \cdot \sqrt{n}, \quad i = \overline{1, p}, \quad j = \overline{1, q}$, as standardized residuals.

The lower these values are in absolute value, the greater the link between the two modalities. The sign indicates the meaning of the link. If the deviation is positive the link is a direct one, the frequency is really high, the two modalities occur simultaneously in a large number of individuals. If the deviation is negative, the two modalities are simultaneously recorded in a few individuals.
Tables 6 and 7 present pairs of modalities with strong direct link. Deviations with values above the mean of the positive deviations were considered.

Table 6. Ethnicity-religion relationship by standardized residuals

Ethnicity	Religion	Ethnicity	Religion
Romanian	Orthodox	Croats	Roman_Catholic

Hungarian	Reformed, Roman_Catholic, Unitarian, Evangelical_Lutheran	Greeks	Old_Calendarists
Romany	Pentecostal	Italians	Roman_Catholic
Ukrainians	Pentecostal	Jews	Judaism
Germans	Evangelical_Augustan, Evangelical_Lutheran, Roman_Catholic	Czechs	Roman_Catholic
Turks	Islam	Poles	Roman_Catholic
Russians_ Lippovans	Old_Calendarists	Chinese	Other_Religion
Tatars	Islam	Armenians	Armenian_Apostolic
Serbs	Serbian_Orthodox	Csango	Roman_Catholic
Slovaks	Evangelical_Lutheran, Roman_Catholic	Macedonians	Islam
Bulgarians	Roman_Catholic	Another	Islam,Other_Religion

Table 7. Ethnicity-language relationship by standardized residuals

Ethnicity	**Language**	**Ethnicity**	**Language**
Romanian	Romanian	Croats	Croatian
Hungarian	Hungarian	Greeks	Greek
Romany	Romani	Italians	Italian
Ukrainians	Ukrainian	Jews	Yiddish
Germans	German	Czechs	Czech
Turks	Turkish	Poles	Polish
Russians_Lippovans	Russian	Chinese	Chinese
Tatars	Tatar	Armenians	Armenian
Serbs	Serbian	Csango	Other_Language
Slovaks	Slovak	Macedonians	Aromanian
Bulgarians	Bulgarian	Another	Other_Language

These tables highlight less obvious associations. Thus, considering the relationship between ethnicity and religions it can be mentioned the relationship between Romany population and Pentecostal religion, Ukrainians and Pentecostals, Bulgarians and Roman Catholic or Macedonians and Islam. The results highlight

multiple significant associations, such as those of the Hungarian population, Germans or Slovaks.

4. Analysis of the differences among ethnic, religious and linguistic diversity at LAU2

Based on relative and marginal frequencies, line and column profile matrices are calculated as follows: $L = \begin{bmatrix} \frac{f_{11}}{f_{1.}} & \frac{f_{12}}{f_{1.}} & \dots & \frac{f_{1q}}{f_{1.}} \\ \frac{f_{21}}{f_{2.}} & \frac{f_{22}}{f_{2.}} & \dots & \frac{f_{2q}}{f_{2.}} \\ \dots & & & \\ \frac{f_{p1}}{f_{p.}} & \frac{f_{p2}}{f_{p.}} & \dots & \frac{f_{pq}}{f_{p.}} \end{bmatrix}$, $C = \begin{bmatrix} \frac{f_{11}}{f_{.1}} & \frac{f_{21}}{f_{.1}} & \dots & \frac{f_{p1}}{f_{.1}} \\ \frac{f_{12}}{f_{.2}} & \frac{f_{22}}{f_{.2}} & \dots & \frac{f_{p2}}{f_{.2}} \\ \dots & & & \\ \frac{f_{1q}}{f_{.q}} & \frac{f_{2q}}{f_{.q}} & \dots & \frac{f_{pq}}{f_{.q}} \end{bmatrix}$.

The sum of the elements of a line in these arrays is 1. Any element L_{ij} represents the share of the profile j in the profile i. For instance, the share of Orthodox in Romanians or the share of Catholics in Hungarians.

Using these matrices, the differences between the observed values and the expected values can be obtained based on the relationship between the two expected values for each administrative unit. If X is the observations table with the share of ethnicity at administrative unit, and Y is the observations table with the share of religions, the expected values can be calculated as follows: $X^e = Y \cdot C$, $Y^e = X \cdot L$. Differences between values will be: $D^X = X - X^e$, $D^Y = Y - Y^e$. So the expected values are those generated by the ethnic structure of each administrative unit for both religion and mother tongue, based on the links between ethnicity and religion / mother tongue synthesized in profile matrices.

Absolute differences will be great for communities with atypical relationships between ethnicity and religion / mother tongue. The values size of the units has a great influence of the differences. Large ethnic, religious or linguistic communities generate larger differences between observed and expected values at the administrative units. The D^X and D^Y matrices can be standardized in order to cancel this influence. Sums of absolute differences for an administrative unit are calculated as follows: $S_k^X = \sum_{i=1}^{p} \left| D_{ki}^X \right|$, $S_k^Y = \sum_{i=1}^{q} \left| D_{ki}^Y \right|$. In the figure 3 are presented maps with differences between the observed and expected values for the religious

profile of units based on relationship between ethnicity and religion. Differences are calculated both: in standardized and non-standardized form.

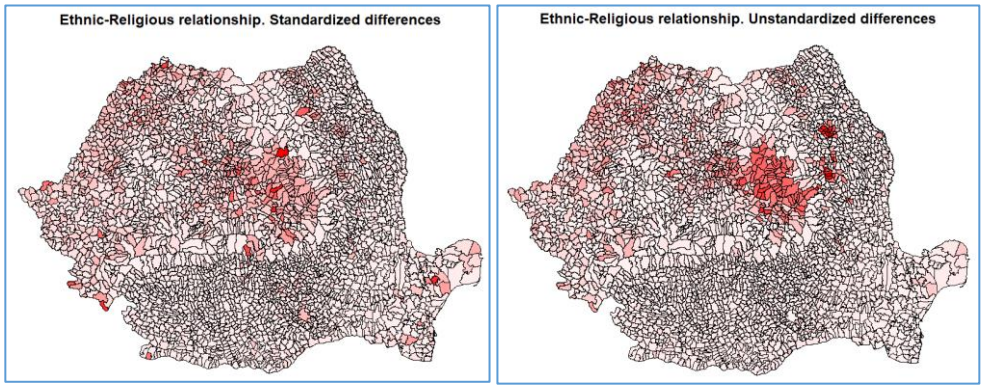

Figure 3. Ethnic-Religious standardized and unstandardized differences

Looking at the map of non-standard differences it can be noticed that the biggest differences are mainly in the Catholic and Reformed communities from the south-east and north-west of Transylvania and Catholic communities from Moldova. These are caused by the diversity of the religious options of the Hungarian and Romany ethnic groups from the mentioned areas and the Romanian ethnic population from some regions of Moldova. Figure 4 shows graphically the religious differences caused by the ethnic structure in a Catholic community from Moldova and a community with ethnic and religious diversity from the south-east of Transylvania. In the first situation the differences are generated because of the entire Romanian community is Catholic, although at the national level the Romanians are Orthodox (93%). In the second situation the data suggest that a large part of the Romanian, Hungarian and Romany ethnic communities are Pentecostal and Unitarians, atypical situation for the three communities.

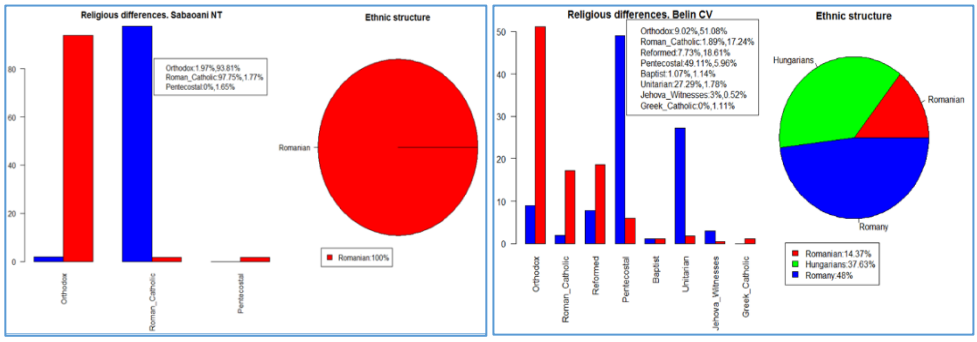

Figure 4. Ethnic-religious differences in Sabauani and Belin villages (cities)

On the map of standardized differences, particular situations are also highlighted in small religious communities. Figure 5 illustrates two such situations. In the first situation the difference is caused by the confession of the Jewish community that is different from the mosaic one, atypical situation. In the second situation there is a great difference between the existing proportion of the Greek Catholics and the one generated by the ethnic structure. The Greek-Catholic population is over 82% of Romanian ethnicity.

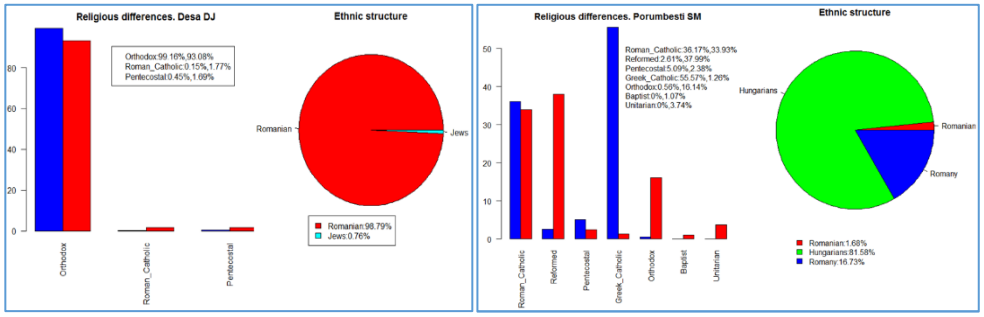

Figure 5.Ethnic-religious differences in Dessa and Porumbesti villages (cities)

The maps of differences between ethnicity and mother tongue are presented in figure 6.

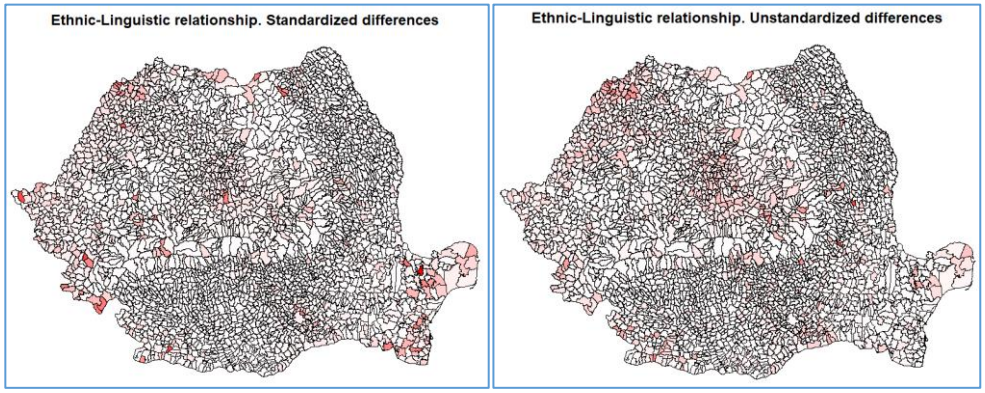

Figure 6. **Ethnic-Linguistic standardized and unstandardized differences**

Differences are lower than those from ethnic-religious case. Religious diversity is greater than linguistic diversity. The high absolute differences occur in the north-west of Romania, in the central area and sporadically in the rest of the territory. In the north-west are mainly caused by the usage of the Hungarian language by the German and Romany communities. In the figure 7 it can be observed how

Hungarian language is used even by German, Romany and Romanian communities. Although the share of Hungarians is only 24.72%, the Hungarian language is used as the mother tongue by over 54% of the population. In the central area and in the rest of the territory, the differences are caused by Romany communities that speak Romani as their mother tongue, given that the fact at the level of the whole country only one third of Romany use Romany language (figure7).

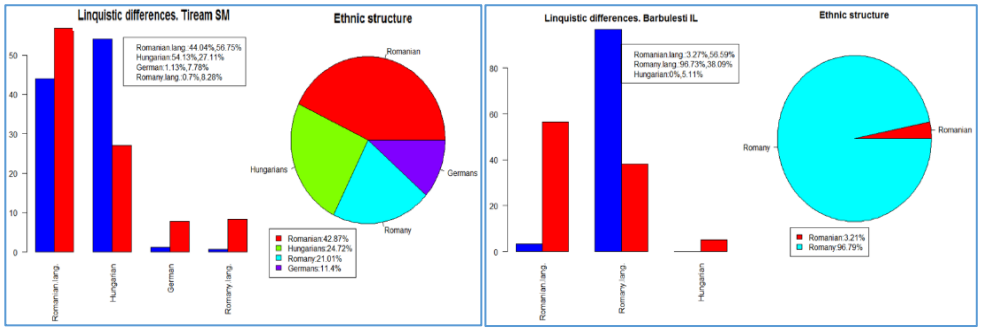

Figure 7. Ethnic-religious differences in Tiream and Barbulesti villages (cities)

The map of standardized differences also highlights the differences generated by the minority languages with a very low share at the national level - Slovak, Polish, Greek, Bulgarian and so on. Figure 8 illustrates two such situations for Croatian and Greek minorities. In both situations, minorities use their own mother tongues, although at national level, important percentages of these minorities use Romanian (for instance Greeks over 30%).

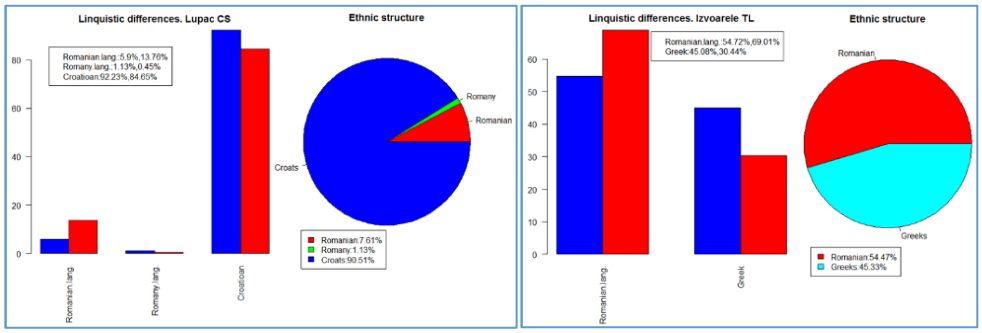

Figure 8. Ethnic-religious differences in Lupac and Izvoarele villages (cities)

5. Usage of diversity indexes in analysis of the differences between ethnic, religious and linguistic diversity

The diversity problem usually arises in biology context in biodiversity studies. There are studies such as [17], [2] which use the biodiversity indexes for analysis of differences between ethnic, religious and linguistic diversity in human communities. There are three main methods used for determining the value of the diversity index: Shannon[10], Simpson and Inverse Simpson[11]. All three methods are implemented by the *diversity* function in the *vegan* R package. The formulas for these indices are:

Shannon-Weaver: $H = -\sum_{i=1}^{n} p_i \log_b p_i$

Simpson: $S1 = 1-\sum_{i=1}^{n} p_i^2$

Inverse Simpson: $S2 = \frac{1}{\sum_{i=1}^{n} p_i^2}$,

where n is the number of species, p_i is the weight of community i (so that $\sum_{i=1}^{n} p_i = 1$) and b is the logarithm base. The natural logarithm ($b=e$) is the default choice for these functions, but b=2 according to information theory fundamentals. The *Simpson* and *Shannon* indices are more sensible to sparse communities compared to the *Inverse Simpson* index. The resulting values are uniformly scaled in the [0,1] interval, where 0 represents perfect uniformity (only one community is present) and 1 represents heterogeneity or diversity. Greater diversity means that the communities are more numerous and with similar counts.A similar computing formula known as Linguistic Diversity Index (LDI) or as the related Index of Linguistic Diversity (ILD) is used to determine linguistic diversity in[5] and [13].
The Simpson index variation for ethnic and religious variations is presented on the map in figure 9. Negative values correspond to the units where religious diversity is lower than ethnic diversity and positive values correspond to the units with greater religious diversity. The map shows greater religious diversity in some communities from Transylvania and Moldavia where the dominant ethnic communities follow different religions. For example, there are Moldavian communities where the Romanians are divided between the orthodox and catholic churches. In the north-west of Transylvania there are Romanian communities divided between Orthodoxism and Greco Catholicism. In the center and south-east of Transylvania the Magyar community is split in Reformed, Catholics and

Unitarians. One such example is presented in figure 10 that shows the ethnic and religious structure in the Călinești commune in Maramureș county in northern Transylvania. In this place we have 7 religions but a single ethnic community. Most of the diversity gap is explained by the division of the Romanian community between Orthodoxism and Greco Catholicism. Table 8 lists the top 5 communities sorted by ethnic-religious diversity differences.

Ethnic diversity greater than religious diversity is generally found in mainly Orthodox communities composed of both Romanians and Romany members.

The discrepancies between ethnic and linguistic diversities are much smaller. The map shown in figure 9 shows this. The same map scale is used for both ethnic-religious and ethnic-linguistic diversity difference maps in order to make them comparable. Table 9 lists the top 5 communities sorted by ethnic-linguistic diversity differences.

There are some typical regional manifestations for the ethnic-linguistic diversity differences:

- Communities where the usage of Hungarian language exceeds the ethnic Hungarian community and extends into Romanian and Romany communities. This happens in some Transylvanian localities. A typical example is Deva (Figure 11) where Hungarian is native language for 7.66% of the population, but the Hungarian ethnics represent under one percent of the population.
- Communities where the Romanian language usage extends to the Romany ethnic community. Examples of such communities can be found across the country.

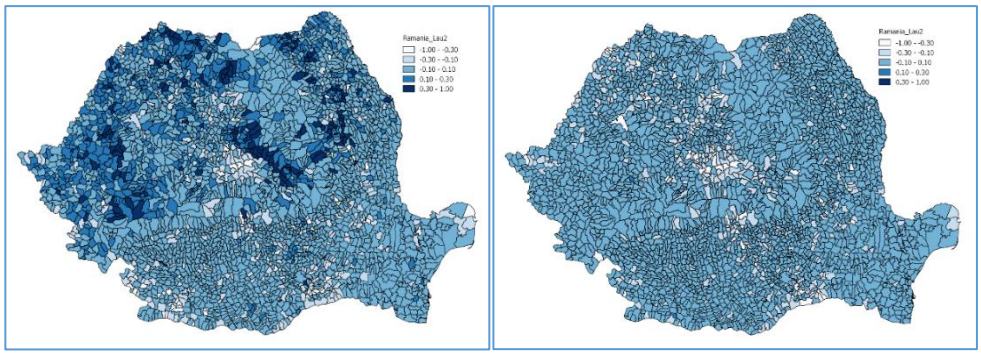

Figure 9. **Ethnic-Religious and ethnic-linguistic diversity differences**

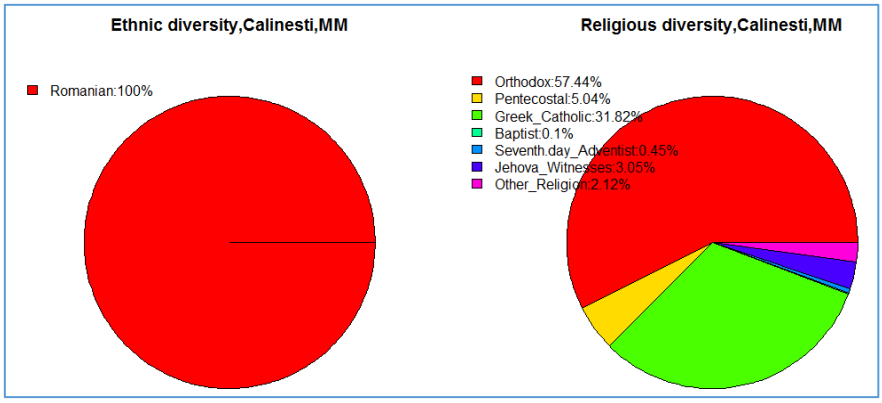

Figure 10. Ethnic and religious structure in the Calinești city, Maramureș County

Table 8. First five cities about ethnic-religious diversity differences. Simpson Index

City/County	Ethnic structure	Religious structure
Galesti/MS Index: 0.669082	Romanian 2.5%; Hungarians 95.85%; Romany 1.65%	Orthodox 2.76%; Roman Catholic 20.45%; Reformed 34.61%; Greek Catholic 0.18%; Baptist 0.26%; Seventh day Adventist 0.77%; Unitarian 28.32%; Jehova Witnesses 5.44%; Romanian Evangelical 0.15%; Without religion 7.06%
Arcus/CV Index: 0.636531	Romanian 3.19%; Hungarians 96.28%; Romany 0.53%	Orthodox 3.59%; Roman Catholic 28.57%; Reformed 30.03%; Seventh day Adventist 1.79%; Unitarian 34.35%; Jehova Witnesses 0.27%; Evangelical Lutheran 0.27%; Other Religion 0.53%; Without religion 0.6%
Baraolt/CV Index: 0.579	Romanian 2.9%; Hungarians 95.94%; Romany 1.06%; Germans 0.09%	Orthodox 2.54%; Roman Catholic 29.23%; Reformed 47.83%; Pentecostal 0.73%; Greek Catholic 0.06%; Baptist 0.99%; Seventh day Adventist 0.04%; Unitarian 16.51%; Jehova Witnesses 0.93%; Evangelical Lutheran 0.24%; Romanian Evangelical 0.04%; Evangelical Augustan 0.05%; Other Religion 0.11%; Without religion 0.71%

Horgesti/BC Index: 0.569867	Romanian 99.84%; Romany 0.16%	Orthodox 59.38%; Roman Catholic 21.85%; Pentecostal 16.1%; Baptist 2.41%; Jehova Witnesses 0.16%; Evangelical 0.09%
Calinesti/MM Index: 0.564903	Romanian 100%	Orthodox 57.44%; Pentecostal 5.04%; Greek Catholic 31.82%; Baptist 0.1%; Seventh day Adventist 0.45%; Jehova Witnesses 3.05%; Other Religion 2.12%

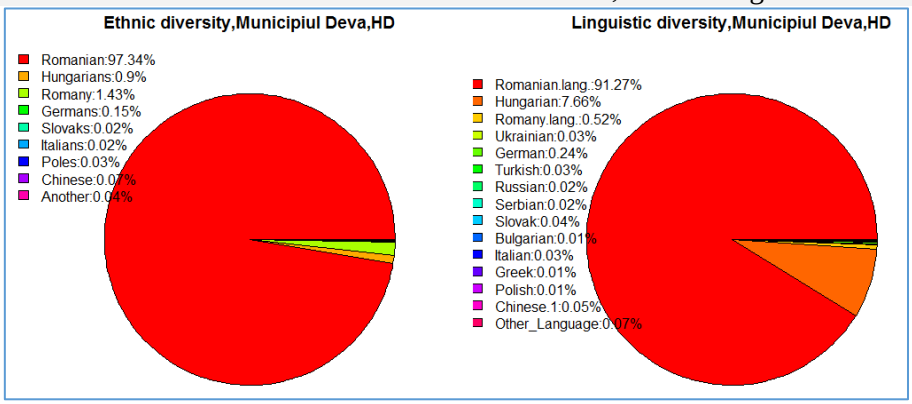

Figure 11. Ethnic and linguistic structure in the Deva city, Hunedoara county

Table 9. First five cities about ethnic-linguistic diversity differences

City/County	Ethnic structure	Linguistic structure
Carcaliu/TL Index: 0.138733	Romanian 9.46%; Russians Lippovans 90.54%	Romanian lang. 19.18%; Russian 80.82%
Deva/HD Index: 0.108893	Romanian 97.34%; Hungarians 0.9%; Romany 1.43%; Germans 0.15%; Slovaks 0.02%; Italians 0.02%; Poles 0.03%; Chinese 0.07%; Another 0.04%	Romanian lang. 91.27%; Hungarian 7.66%; Romany lang. 0.52%; Ukrainian 0.03%; German 0.24%; Turkish 0.03%; Russian 0.02%; Serbian 0.02%; Slovak 0.04%; Bulgarian 0.01%; Italian 0.03%; Greek 0.01%; Polish 0.01%; Chinese 0.05%; Other Language 0.07%
Cojasca/DB Index: 0.106666	Romanian 22.08%; Romany 77.88%; Turks 0.04%	Romanian lang. 65.6%; Romany lang. 34.4%
Gangiova/DJ Index: 0.104209	Romanian 100%	Romanian lang. 94.49%; Romany lang. 5.51%
Rastolita/MS Index: 0.073444	Romanian 89.6%; Hungarians 2.65%; Romany 7.74%	Romanian lang. 84.4%; Hungarian 15.4%; Romany lang. 0.2%

Conclusions

Application of the statistical methods has shown that in Romania the relationship between ethnicity, religion and mother tongue is very strong and consistent. There is a specific pattern for the relationship between these elements. This is explained by the fact that, in Romania, there are traditional ethnic, religious and linguistic communities with a long history behind them, unlike the western European countries that have many communities formed in the last 30 years through immigration. In Romania most of the communities have kept their cultural identity. This paper identifies the cases where the cultural identity was lost by analyzing the exceptions from the typically strong ethnicity – religion – mother tongue relationship. Most of those exceptions are identified inside the Romany communities. An important aspect is the fact that the identity changes are not always geared towards the majority culture (Romanian ethnicity, Orthodox faith and Romanian language). An example for this (shown in section 4) is the case of the loss of linguistic identity inside the German communities located in the north-vest regions of Romania by switching to the Hungarian language. The results presented in this study can by refined further by extracting zone with specific patterns and identifying local exceptions inside those regions.

REFERENCES

[1]**Alesina, A., La Ferrara, E. (2005),** ***Ethnic diversity and Economic Performance*****;** Journal of Economic Literature 43(3): 762-800;

[2] **Allen, J.P. (2005),** ***Ethnic Geography Dynamics: Clues from Los Angeles*****;** Association of Pacific Coast Geographers Yearbook, 67, 97-116;

[3]**Dârdală, M., Furtună, T.F., Reveiu, A.(2016),** ***Integrating R Scripts in GIS*****;** Vol. Education, Research & Business Technologies; Proceedings of the 15th International Conference on Informatics in Economy (IE 2016), Cluj-Napoca, Romania, June 02 –05, pp. 250-256;

[4] **Gower, J.C.(1989),** ***Generalized Canonical Analysis***, in R. Coppi and S. Bolasco (eds.), Multiway Data Analysis, pp.221-232, Amsterdam;

[5] **Joseph H.G. (1956),** ***The Measurement of Linguistic Diversity***; JSTOR, Vol. 32, No. 1, pp. 109-115;

[6]**Kettenring, R.J.(1971),** ***Canonical Analysis of Several Sets of Variables***; Biometrica, vol. 58, No. 3, pp. 433-451;

[7]**Marcon, E.(2015),** ***Mesures de la Biodiversite***; Master, Kourou, France;

[8] **Putnam, R.E. (2007),** ***Pluribus Unum: Diversity and Community in the Twenty-First Century*****;** The 2006 Johan Skytte Prize Lecture. Scandinavian Political Studies 30(2): 137-174;

[9] **Schaeffer, M.(2014),** ***Ethnic Diversity and Social Cohesion. Immigration, Ethnic Fractionalization and Potentials for Civic Action***; Farnham (UK), Burlington (USA): Ashgate;
[10]**Shannon, C.E., Weaver, W.(1963),** ***The Mathematical Theory of Communication****;* University of Illinois Press;
[11]**Simpson, E.H. (1949),** ***Measurement of Diversity***. Nature 163(4148): 688;
[12]**Snedecor, G.W., Cochran, W. G. (1989),** ***Statistical Methods***; Eighth Edition, Iowa State University Press;
[13] **Steele, J. (2008),** ***Population Structure and Diversity Indices***; Genetics and Human Prehistory, McDonald Institute for Archaeological Research. Chapter 18: 187-191;
[14]**Stolle D., Soroka S., Johnston R(2008),** ***When Does Diversity Erode Trust? Neighbourhood Diversity***; Interpersonal Trust and the Mediating Effect of Social Interactions, Political Studies 56(1): 57-75;
[15] **Sturgis, P., Brunton-Smith, I., Kuha, J., Jackson, J. (2014),** ***Ethnic Diversity, Segregation and the Social Cohesion of Neighbourhoods in London***; Ethnic and Racial Studies, Vol. 37, No. 8, 1286–1309;
[16]**Tenenhaus, A.,Tenenhaus, M. (2014),** ***Regularized Generalized Canonical Correlation Analysis for Multiblock or Multigroup Data Analysis***; European Journal of Operational Research, vol. 238, 391-403;
[17] **White, M.J. (1986),** ***Segregation and Diversity Measures in Population Distribution***; Population Index, 52, 198-221;
[18] **Whittaker, R.H. (1972),** ***Evolution and Measurement of Species Diversity***; Taxon, 21(2/3): 213-251;
[19]http://www.recensamantromania.ro/rezultate-2/.