ANALYZING THE CULTURE CONSUMERS AT THE TERRITORIAL LEVEL BY THE PRINCIPAL COMPONENT METHOD

Marioara Mirea[1*], Kamer-Ainur Aivaz[2]
[1) 2)] *"Ovidius" University of Constanta, Romania, Faculty of Economic Sciences*

Abstract
The main objective of this paper concerns the interaction between four indicators expressing cultural consumption and two significant indicators expressing the consumption availability that satisfies the population's spiritual needs, in each Romanian county and in Bucharest, in 2014.
The variables introduced in the analysis are "Active readers in libraries", "Spectators at cinemas", "Spectators and auditors to artistic performances", "Visitors to museums and public collections", "Unemployment" and "Average monthly net nominal earnings". The results triggered by the statistical method "Principal components analysis (PCA)" revealed significant correlations between the variables taken into consideration and highlighted the similarities and differences between the analyzed statistical units.

Keywords
Active readers in libraries, spectators and auditors to artistic performances, average net nominal earnings, principal component analysis.

JEL Classification
I20, J30, Z10, C10, C38

Introduction

The main research objective is represented by the analysis of book consumers, performances at cinemas, artistic performances, museums and public collections, depending on the population's cash availability and the indicators "Unemployment" and "Average monthly net nominal earnings".
For the purpose of this analysis, we compiled the latest statistical data on the cultural consumption in all 41 Romanian counties and Bucharest, data provided by the National Institute of Statistics - Romania (NIS), the reference period being the calendar year 2014.
The variables considered in this analysis are "Active readers in libraries", "Spectators at cinemas", "Spectators and auditors to artistic performances", "Visitors to museums and public collections", "Unemployment" and "Average monthly net nominal earnings", which

* Corresponding author, **Marioara Mirea** - mm_mirea@yahoo.com

characterize the population's cultural consumption preferences per consumption profiles.
The study assesses the interaction between these variables by applying the statistical method known as "Principal component analysis" (PCA); data processing, testing the significance of indicators and the graphical representations were performed by the SPSS statistical software. (Field, 2009)
Since the aim of this paper is to analyze the way in which the population consumes cultural products, it is necessary to make some methodological explanations regarding the content of indicators according to the defining elements provided by the National Institute of Statistics - Romania (NIS).
The indicator "Active readers in libraries" includes the individuals who consulted in the library or borrowed at least once a year a book or a publication belonging to a library.
The indicator "Spectators at cinemas" includes the cinephile audience that was statistically recorded, based on the tickets sold at movie theaters.
"Spectators and auditors to artistic performances" records the number of the spectators who attended a theater performance or concert, based on the tickets sold.
According to statistical data, the indicator with the highest value, i.e. "Visitors to museums and public collections", includes the number of the visitors to a museum or public collection, based on the tickets sold and the participants to the "Night of Museums".

Data and results

After processing the data, we obtained the statistical indicators calculated for the analyzed variables and the graphical representation of these points in the system of factorial axes. It shows the distribution of the 6 variables for each Romanian county and for Bucharest, in 2014.
Taken together, the information on the cultural consumption per consumer profiles shows that the indicator "Visitors to museums and public collections" ranks first, with 34.94% of all culture users, followed by the category "Spectators in cinemas", with 32.63%. The audience devoted to certain artistic performances recorded a percentage of 20.14% and the lowest percentage belongs to the reading of books, i.e. the category "Active readers in libraries", with 12.29%.
At territorial level, the statistics indicate the following situation:

- For the category "Active readers in libraries", the largest number of readers is registered in the following counties: Bucharest Municipality (346,188), Cluj (176,682), Iasi (169,914), Valcea (145,794), Timis (133,499), Suceava (129,165), Prahova (128,267), Arges (127,855); the lowest number is registered in Ilfov (24,122), Giurgiu (35,243), Calarasi (39,014), Covasna (40,267), Caras-Severin (43,722), Salaj (43,895);
- For the category "Spectators in cinemas", the largest number of spectators is registered in the following counties: Bucharest (4,292,000), Cluj (803,000), Constanta (688,000), Timis (545,803), Prahova (464,000), Iasi (397,000), Bihor (367,000); the lowest number is registered in the following counties: Neamt (4,000), Ialomita (5,000), Sibiu (7,000);
- For the category "Spectators and auditors to artistic performances", the largest number of spectators is registered in the following counties: Bucharest Municipality (1,137,510), Olt (800,950), Suceava (621,000), Dolj (308,963), Cluj (260,509), Sibiu (239,066), Timis (228, 987); the lowest number is registered in Teleorman (11,000), Neamt (11,670), Vaslui (16,124), Caras-Severin (18,361);
- For the category "Visits to museums and public collections", the largest number of

spectators is registered in the following counties: Bucharest Municipality (1,462,410), Brasov (1,400,705), Sibiu (882,117), Prahova (698,577); the lowest number is registered in Giurgiu (3,292), Teleorman (5,501), Olt (14,544), Mehedinti (17,847).

It is noteworthy that, in Bucharest Municipality, all indicators, both in terms of the number of spectators and readers, recorded the highest levels, while in the other counties, the situation for each category of spectators reveals some differences, depending on the economic and spiritual features of the respective area.

a. Descriptive statistics indicators (Descriptive Statistics output)

For the category "Active readers in libraries", the average of the readers at country level is 180,002 readers. Thus, apart from Bucharest, all the other counties are situated below this average; from this perspective, the series show a pronounced elongation (kurtosis is 41,189). Basically, 97.6% of the counties are situated below the average.

For the category "Spectators in cinemas", the country-wide average is 480,810 spectators; 4 counties, i.e. Bucharest, Cluj, Constanta and Timis, hover over this medium level. Thus, 97.36% of the counties analyzed are situated below the average.

For the category "Spectators and auditors to artistic performances," the country-wide average is 291,926 spectators; 9 counties (i.e. 21.4%) - Bucharest, Olt, Suceava, Dolj, Cluj, Sibiu, Timis, Mures and Vrancea - recorded a number of viewers above this value.

For the category "Visitors to museums and public collections", the country-wide average is 514,697; 4 counties (9.5%) - Bucharest, Brasov, Sibiu and Prahova - registered values above the average.

It is also noteworthy that, at county level, these indicators show large asymmetries.

b. Correlation matrix

The correlation matrix shows the values of the correlation coefficients for the variables considered in twos. It is a square matrix, symmetrical to the main diagonal (equal to one because a variable is perfectly correlated with itself). The form of the correlation matrix is shown in table no.1.

The analysis of the correlation matrix coefficients allows the assessment of the possibility of applying principal component analysis. The high values of these coefficients (greater than +0.5 or less than -0.5) indicate that there are statistically significant connections between the variables considered (direct connections if the value of these coefficients is positive, reverse connections if the value of these coefficients is negative). In this case, principal component analysis can be applied. The low values of these coefficients show that there are no correlations between statistical variables and, therefore, PCA, whose purpose is to identify these correlations, cannot be applied.

Table no. 1 The correlation matrix of the variables "Active readers in libraries," "Spectators at cinemas", "Spectators and auditors to artistic performances", "Visitors to museums and public collections", "Unemployment" and "Average monthly net nominal earnings"

Correlation Matrix[a]							
		Active readers in libraries	Spectators at cinemas	Spectators and auditors to artistic performances	Visitors to museums and public collections	Unemployment	Average monthly net nominal earnings
Correlation	Active readers in libraries	1.000	.943	.984	.989	-.074	.187
	Spectators at cinemas	.943	1.000	.955	.946	-.199	.395
	Spectators and auditors to artistic performances	.984	.955	1.000	.977	-.081	.245
	Visitors to museums and public collections	.989	.946	.977	1.000	-.118	.232
	Unemployment	-.074	-.199	-.081	-.118	1.000	-.456
	Average monthly net nominal earnings	.187	.395	.245	.232	-.456	1.000

Source: SPSS processing, based on data from the National Institute of Statistics

A correlation matrix feature is that the number of correlation coefficients greatly increases when the number of variables (k) included in the analysis increases, regardless of the statistical collectivity volume. The number of correlation coefficients is k(k-1)/2. For a data table showing the values of six variables, the number of correlation coefficients is 15 (see table no.1). This significant increase in the number of correlation coefficients highlights the possibility to interpret the connections between variables only by analyzing the values presented in the correlation Matrix.

c. Calculating the distance χ^2

χ^2 statistics is used in order to test the hypothesis of the independence between the variables studied. For this purpose, the following hypotheses are made:

- hypothesis H_0, which is the hypothesis of independence (the correlation matrix is an identity matrix), which acknowledges that there are no statistical connections between variables;
- hypothesis H_1 is the hypothesis of dependence, which acknowledges that there are connections between statistical variables. (Dimitrios, Stephen, 2011)
- In order to test these hypotheses, the SPSS software provides (KMO output and Bartlett's Test) both the calculated value of the test statistics ($\chi^2_{calculated}$= 409,275) and the probability value associated to the calculated test statistics (Sig.)

Table no. 2 The value of test statistics χ^2

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.801
Bartlett's Test of Sphericity	Approx. Chi-Square	409.275
	df	15
	Sig.	.000

Source: SPSS processing, based on data from the National Institute of Statistics

A value Sig. = 0.000 <0.05 associated with the calculated value of the test statistics χ^2 shows that the hypothesis H_0 is rejected and the hypothesis H_1 is accepted. It can thus guarantee, with a 95% probability, that there are statistically significant connections between statistical variables. In this situation, the factorial analysis can be applied to the data considered.

The simultaneous analysis of the results triggered by testing the hypothesis of independence, using the test statistics χ^2, and the determinant value of the correlation matrix allows identifying the properties of this matrix of interest to PCA.

Identifying the existence of the connections between variables is facilitated by calculating the Kaiser-Meyer-Olkin statistics (KMO), Measure of Sampling Adequacy. The values of the KMO statistics can range between 0 and 1. A value greater than 0.5 indicates that there are significant connections between statistical variables; therefore, PCA can be applied. (Pintilescu, 2007). Based on the results outlined in table no.4, the KMO value of 0.801 indicates that there are statistical connections between the analyzed variables, PCA thus providing a good solution.

d. The variance of variables (Communalities output)

The standardization of variables yields new variables of zero mean and variance one. The variances of statistical variables are presented in the Communalities output, as follows:

Table no. 3 The variance of statistical variables²

Communalities		
	Initial	Extraction
Active readers in libraries	1.000	.987
Spectators at cinemas	1.000	.967
Spectators and auditors to artistic performances	1.000	.985
Visitors to museums and public collections	1.000	.980
Unemployment	1.000	.746
Average monthly net nominal earnings	1.000	.725

Source: SPSS processing, based on data from the National Institute of Statistics

The variance values after extracting the factors are calculated based on the results of the Component Matrix output (table no.5). For example, for the variable "Active readers in libraries", we get $\sigma_i = 0{,}975^2 + 0{,}194^2 = 0{,}987$. The high values of the variables' variance after the extraction of factors (Extraction column) reveal that those variables should not be removed from the proper analysis because they are connected to the factorial axes. (Everitt, Dunn, 2001; Pintilescu, 2007)

e. The eigenvalues λ_k, associated to each factorial axis and the variance explained by each factorial axis (Total Variance Explained output)

The eigenvalues of the correlations matrix are shown in the *Total Variance Explained output, column Initial Eigenvalues* (table no.4).

Table no. 4 Eigenvalues of the correlations matrix and the variance explained by the factorial axes

Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.023	67.048	67.048	4.023	67.048	67.048
2	1.368	22.801	89.850	1.368	22.801	89.850
3	.536	8.941	98.791			
4	.043	.722	99.513			
5	.021	.350	99.864			
6	.008	.136	100.000			

Source: SPSS processing, based on data from the National Institute of Statistics

In the output above, the eigenvalues of the correlation matrix are λ_1=4,023, λ_2=1,368, λ_3=0,536, λ_4=0,043, λ_5=0,021 and λ_6=0,008. As already mentioned, the eigenvalues correspond to the inertia explained by the factorial axes. Their sum is the total inertia of the cloud of points equal to the number of the statistical variables from the original data table, i.e. the sum of the main diagonal elements of the correlation matrix: $\sum_{k=1}^{K} \lambda_k = 6$

The software displays these values as absolute (column *Total*) or relative (column *% of Variance*) or as a percentage of the total inertia, as well as cumulative relative values (column *Cumulative %*). (Kachigan, 1982; Pintilescu, 2007)

Thus, the first factorial axis explains 67.048% of the total variance of the cloud of points; the first two factorial axes explain together 89.85% of the total variance; the first three factorial axes explain 98.791% and so on. According to Benzecri criterion, which involves choosing the number of the axes explaining over 70% of the total variance of the cloud of points, for the purpose of our study, we will choose the first two factorial axes. (Benzecri, 1992)

f. The coordinates of the variables on the factorial axes (Component Matrix output)

The coordinates of the variables on the factorial axes show the values of the correlation coefficients between variables and the respective factorial axis.

Table no. 5 The coordinates of the variables on the first two factorial axes

Component Matrix[a]		
	Component	
	1	2
Active readers in libraries	.975	.194
Spectators at cinemas	.983	-.016
Spectators and auditors to artistic performances	.980	.155
Visitors to museums and public collections	.980	.140
Unemployment	-.209	.838
Average monthly net nominal earnings	.375	-.765

Source: SPSS processing, based on data from the National Institute of Statistics

The values in table no.5 show the position of variables on factorial axes. For example, the variable "Active readers in libraries" has a high positive coordinate (close to 1) on the first factorial axis (0.975) and a low positive coordinate on the second factorial axis (0.194). This shows that the variable will be graphed in the positive quadrant on both factorial axes. The variable "Spectators at cinemas" has a positive coordinate on the first factorial axis (0.983) and a negative coordinate on the second factorial axis (-0.016). The high values of the variables' coordinate on the factorial axes show that those variables are highly correlated with the respective factorial axis.

The variables' coordinates on the factorial axes are the linear equation coefficients of the connections between variables. For the data in table no.5, the first factorial axis is a new variable defined by the linear combination of the initial variables, of the form:

F_1=0, "Active readers in libraries" +0.983 "Spectators at cinemas" +0.980 "Spectators and auditors to artistic performances" +0.980 "Visitors to museums and public collections" -0.209 "Unemployment" +0.375 "Average monthly net nominal earnings".

In order to identify the variables that explain the second factorial axis, we select those variables from table no.5 (column Component 2) that have higher coordinate values. It is noteworthy that the formation of the second factorial axis is explained, for the example, only by "Unemployment" and "Average monthly net nominal earnings".

g. The contribution of variable X_j to the inertia of axis k (Component Score Coefficient Matrix output)

The high values of the contributions reveal the significant importance of the respective variable in differentiating the statistical units considered (Spircu, L., 2005). Thus, (according to table no.6), all variables contribute to the formation of the first factorial axis, and, as far as the second axis is concerned, the only variable that does not contribute is "Spectators at cinemas".

Table no. 6 The contribution of variables to the inertia of the first two factorial axes

Component Score Coefficient Matrix		
	Component	
	1	2
Active readers in libraries	.242	.141
Spectators at cinemas	.244	-.011
Spectators and auditors to artistic performances	.244	.113
Visitors to museums and public collections	.244	.102
Unemployment	-.052	.612
Average monthly net nominal earnings	.093	-.559

Source: SPSS processing, based on data from the National Institute of Statistics

Conclusions

The graphical representation, which is actually a review of the previous results, allows viewing the position of variables in the factorial axes system, identifying the direction and intensity of the relationship between variables. The graphs highlight easier the position of variables in the factorial axes system and identify the direction and intensity of the connections between the analyzed variables:

- as far as the direction of the connection between variables is concerned, it can be considered that there is a direct connection between the variables represented on the same side of a factorial axis.
- as far as the intensity of the connections between variables is concerned, it can be considered that there is a strong connection between the variables represented close to the correlation circle.

As far as the direction is concerned, it can be considered that there is a direct connection between the variables represented on the same side (positive or negative) of a factorial axis. These variables are positively correlated with each other and have the same sign of the coordinate on the respective factorial axis. Therefore, on the first factorial axis, there is a reverse connection between the points "Active readers in libraries," "Spectators at cinemas", "Spectators and auditors to artistic performances", "Visitors to museums and public collections" and "Average monthly net nominal earnings" on the one side, and "Unemployment", on the other side. There is a direct connection between the points "Active readers in libraries," "Spectators at cinemas", "Spectators and auditors to artistic performances", "Visits to museums and public collections" and "Average monthly net nominal earnings", variables situated on the right side of the graph.

On the second factorial axis, there are significant differences between "Active readers in libraries", "Spectators and auditors to artistic performances", "Visitors to museums and public collections" on one side, and "Spectators at cinemas" on the other.

As far as the counties are concerned, the factorial axis 1 (the vertical one) indicates that there are vast differences between the Municipality of Bucharest (28), Timis (38), Vrancea (42) on one side, and Alba (1), Mures (29), Tulcea (39) on the other side, regarding the structure of the participants in the act of culture in all its forms.

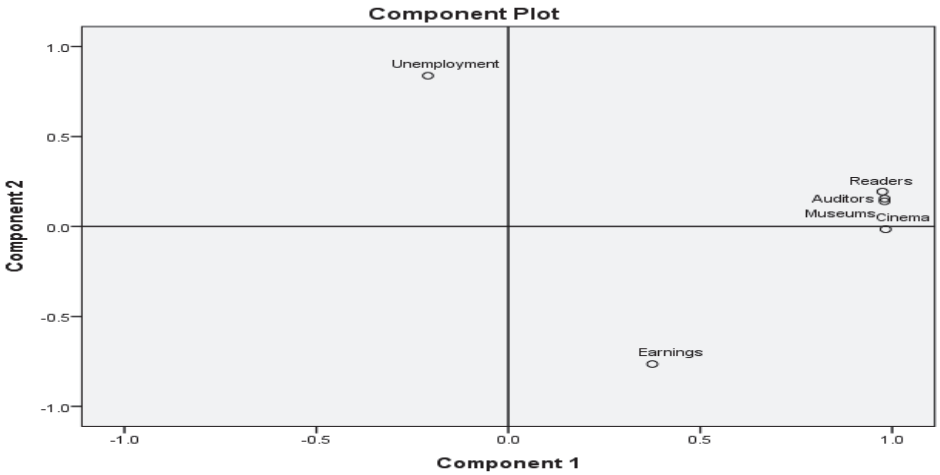

Figure no.1 The graphical representation of the variables' position in the system of the first two factorial axes

Source: Processing based on data from the National Institute of Statistics

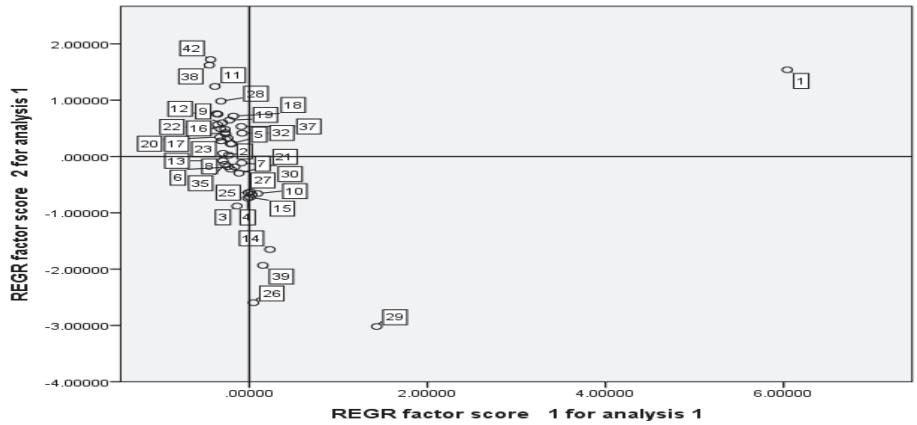

Figure no.2 The graphical representation of the territorial units (counties) on the first two factorial axes

Source: Processing based on data from the National Institute of Statistics

References

Benzecri, J.P., 1992. *Correspondence Analysis Handbook*. New-York.
Dimitrios, A.,Stephen, G.H., 2011. *Applied Econometrics.* Palgrave Macmillan
Everitt, B., Dunn, G., 2001. *Applied Multivariate Data Analysis.* Arnold, London.
Field, A., 2009. *Discovering Statistics Using SPSS.* Sage Publications Ltd., London.

Institutul Naţional de Statistică. [online] Available at: < http://statistici.insse.ro/> [Accessed 20 February 2016].
Kachigan, S., 1982. *Multivariate statistical analysis.* Radius Press, New York.
Pintilescu, C., 2007. *Analiză statistică multivariată* (*Multivariate analysis)*. Editura Universităţii Alexandru Ioan Cuza, Iaşi.
Spircu, L., 2005. *Analiza datelor-Aplicaţii economice (Economic Data Analysis Applications)*. Editura ASE, Bucureşti