ANALELE ŞTIINŢIFICE ALE UNIVERSITĂŢII „ALEXANDRU IOAN CUZA" DIN IAŞI
Tomul LIV Ştiinţe Economice 2007

DISCRIMINANT ANALYSIS IN THE STUDY OF ROMANIAN REGIONAL ECONOMIC DEVELOPMENT

ELISABETA JABA*
DĂNUŢ VASILE JEMNA*
DANIELA VIORICĂ*
CHRISTIANA BRIGITTE BALAN*

Abstract

In this paper we propose using the discriminant analysis for the identification of a typology for the Romanian counties by the economic development level. In this purpose, we used a set of variables that characterize the economic and social development. The treatment of the data is done with the SPSS software. The results obtained in this paper can be used as arguments in making decisions regarding the harmonization of the Romanian regions and the allocation of the investments in appropriate counties and regions.

Key words: European integration process, regional economic development, discriminant analysis.

1 Introduction

In the classical model [], the evaluation of the counties' development level is based on either using a system of indicators, or one synthetic indicator.

We consider that this approach has disadvantages because it takes into account the level of the indicators and not the relationships between the variables that explain the economic development.

In this paper we propose using the discriminant analysis for the identification of a typology for the Romanian counties by the economic development level. The discriminant analysis presents the advantage to synthesize a set of variables using the discriminant function. Moreover, it expresses the relationship between the variables from the set used to characterize the development level and the discriminant function score.

The statistical observation has been carried out on a set of variables of the development recorded at the level of the 42 counties of Romania, grouped into eight development regions. The county of Bucharest has been excluded from this analysis since its values place it as an outlier.

* Professor, PhD, Department of Statistics, Faculty of Economics and Business Administration, "Alexandru Ioan Cuza" University, Iasi, e-mail: ejaba@uaic.ro
* Associate professor, Department of Statistics, Faculty of Economics and Business Administration, "Alexandru Ioan Cuza" University, Iasi, e-mail: djemna@uaic.ro
* Assistant, Department of Statistics, Faculty of Economics and Business Administration, "Alexandru Ioan Cuza" University, Iasi, e-mail: dana.viorica@gmail.com
* Assistant, Department of Statistics, Faculty of Economics and Business Administration, "Alexandru Ioan Cuza" University, Iasi, e-mail: christiana.brigitte@yahoo.com

The data used in this analysis have been taken from the official European Statistics (EUROSTAT) as well as from the national statistics (National Institute of Statistics). The analysis was carried out for the year 2003, this being the year for which we had the latest regional statistical data.

2 Method

2.1 General Elements

The discriminant analysis is a multivariate statistical method used to estimate the linear relationship between a dependent non-metrical variable having two or more categories and linear combinations of more independent metric variables. The relationship is estimated by the following discriminant function:

$$D = b_1 X_1 + b_2 X_2 + ... + b_n X_n + c$$

where

D = discriminant variable;

X_i=discriminating (independent) variables;

b_i= discriminant coefficients;

c= constant

In our study the independent variables are the variables that characterize the economic and social development level of the Romanian counties. They are the following: X_1-*phys2003* (number of physicians per 1000 inhabitants), X_2 –*urb2003* (urbanization level), X_3-*wage2003* (nominal net mean wages), X_4-*stud2003* (higher education school population), X_5-*work2003* (number of people that have had accidents at work place), X_6-*RD2003* (research and development expenses), X_7-*pop2003* (number of inhabitants), and X_8-*libr2003* (number of libraries).

The discriminant variable by which we divide the counties in groups is the labour productivity, expressed in billions of ROL/employee. This variable takes values from 1 (very low labor productivity) to 4 (very high labour productivity). The four values correspond to the following numerical intervals: 1 for the interval 140-180 million lei/employee, 2 for interval 180-220, 3 for interval 220-260, and 4 for interval 260-300 (see fig. 1).

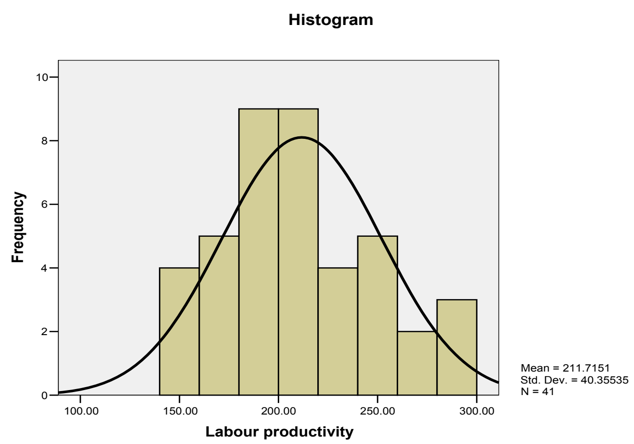

Fig. 1 Distribution of the counties of Romania according to labour productivity in 2003

2.2 Conditions for Applying DA

The application of the discriminant DA implies checking up certain hypotheses regarding:

• normality of multivariate distributions – the predictor variables must have normal multivariate distributions. However, the discriminant analysis is relatively robust even when the multivariate normal distribution are contradicted [Lachenbruch, P. A, 1975]. The dichotomous variables, which very often contradict the hypothesis of multivariate normal distribution, do not affect the discriminant analysis conclusions [Klecka, W. R., 1980].

• homogeneity of variances (homoscedasticity) – within each group, the variance of each independent variable must be the same. That is, the independent variables may have different variances between them, but for the same independent variable the variances and group means must be equal. The absence of variances homogeneity can indicate the presence of outliers in one or several groups.

• absence of multi-co-linearity – if one of the independent variables is strongly correlated with another independent variable, or one of the independent variables is a function (e.g., a sum) of other independent variables, then the tolerance value for that variable will be close to 0 and the matrix will have no unique discriminant solution.

3 Results

3.1 Selection of Discriminating Variables

In order to determine the variables which significantly contribute to the differentiation of groups, we have used test F for Wilks's Lambda. The ANOVA results are given in Table 3.

Table 3. Tests of Equality of Group Means

Variable	Wilks' Lambda	F	df1	df2	Sig.
phys2003	0.785	3.372	3	37	0.029
urb2003	0.765	3.788	3	37	0.018
wage2003	0.604	8.087	3	37	0.000
stud2003	0.813	2.844	3	37	0.051
work2003	0.848	2.203	3	37	0.104
RD2003	0.727	4.634	3	37	0.008
pop2003	0.882	1.647	3	37	0.195
libr2003	0.886	1.581	3	37	0.210

Test F is significant for five variables out of 8 (values of Sig. smaller than 0.05 for the first three variables as well as for the sixth variable, and smaller than 0,1 for the fourth variable). The last two variables, for which Sig. is higher than 0.1, should be eliminated from the model.

3.2 Estimation of Discriminant Function

In our study, the discriminant analysis was carried out for 4 groups of counties by the productivity level and it resulted in 3 discriminant functions and consequently 3 eigenvalues.

The highest eigenvalue (1.818) corresponds to the first discriminant function, which shows that it has the strongest power of discrimination of the three functions. Also, the first function accounts in a ratio of 80.5% for the dispersion of the group means, as compared to the other two functions, which, taken together, account for less than 20% of dispersion (see Table 4).

The canonical correlation coefficient, measuring the relation between the discriminant factorial coordinates and the grouping variable, shows that 64,4%, that is $(0.803)^2$, of the total variance accounts for the differences among the four groups through the first discriminant function. (see Table 4).

Table 4. Eigenvalues

Function	**Eigenvalue**	**% of Variance**	**Cumulative %**	**Canonical Correlation**
1	1.818[a]	80.5	80.5	0.803
2	0.357[a]	15.8	96.3	0.513
3	0.084[a]	3.7	100.0	0.278

The discriminating variables considered in our study are expressed in different units of measure, and consequently the standardized coefficients of the discriminant function were calculated [Jaba E., Grama, A., 2004]. They are given in Table 5.

Table 5. Standardized Canonical Discriminant Function Coefficients

Variable	**Function**		
	1	**2**	**3**
phys2003	0.479	0.360	0.637
urb2003	0.540	0.615	-0.412
wage2003	0.704	-0.044	-0.329
stud2003	-0.252	-0.856	-1.372
work2003	-0.267	0.240	0.389
RD2003	0.712	-0.056	0.457
pop2003	0.526	-0.744	1.541
libr2003	-0.738	1.585	-0.791

The discriminant function coefficients are used for calculating the discriminant score for each case in particular. Taking into account that the first function has the highest discriminating power, we shall focus our attention upon analyzing its results.

Therefore, the first discriminant function is

$Z = 0.479Z_1 + 0.540Z_2 + 0.704Z_3 – 0.252Z_4 -0.267Z_5 + 0.712Z_6 + 0.526Z_7 – 0.738Z_8$

where variables Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7 and Z_8 are standardized X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8 variables.

The size of the coefficients indicates the discriminant power of the predictor variables. Thus it can be seen that the variables *number of libraries* (X_8), *research and development expenses* (X_6), *mean nominal net wages* (X_3), *level of urbanization* (X_2), *number of inhabitants* (X_7), and *number of physicians per inhabitant* (X_1) discriminate best among the four groups.

The structure matrix coefficient indicates the correlation between each predictor variable and the discriminant function. The values of the structure coefficients obtained are presented in Table 6.

Table 6. Structure Matrix

Variable	**Function**		
	1	**2**	**3**
wage2003	0.600	-0.075	0.029
RD2003	0.450	0.030	0.300
phys2003	0.359	0.296	-0.300

Variable	Function		
	1	2	3
libr2003	0.067	0.577	0.114
urb2003	0.321	0.555	-0.351
work2003	0.229	0.473	0.202
pop2003	0.204	0.380	0.274
stud2003	0.332	0.227	-0.370

For the first discriminant function, it can be seen that the correlation coefficients have high values for the first three variables, which means that these variables are most strongly correlated with the first function.

For the second discriminant function, the coefficients show that the next four variables are strongly correlated, while the third discriminant function is correlated with the last variable.

The first discriminant function is correlated with three indicators from three different classes: economic, education and health, groups which would probably characterize best the division of counties according to the degree of social-economic development. A second function also has variables from three classes of indicators: education, demography and health. A classification of the counties by using the scores obtained for this function would sooner reflect the social and demographic condition of the counties.

3.3. Efficiency of Discriminant Function

In our study, based on the discriminant function, 65.9% of the counties have been correctly classified, that is (7+10+5+5)/41 (see Table 7). A case is considered to be classified correctly if, by the discriminant function score, it is included in the group to which it actually belongs.

Table 7. Classification Results (a)

Productivity			Predicted Group Membership				Total
			Very low productivity	Low productivity	High productivity	Very high productivity	
Original	Count	Very low productivity	7	2	0	0	9
		Low productivity	6	10	0	2	18
		High productivity	0	3	5	1	9
		Very high productivity	0	0	0	5	5
	%	Very low productivity	77.8	22.2	0.0	.0	100.0
		Low productivity	33.3	55.6	0.0	11.1	100.0
		High productivity	0.0	33.3	55.6	11.1	100.0
		Very high productivity	0.0	0.0	0.0	100.0	100.0

(a) 65.9% of original grouped cases correctly classified.

Table 8. Classification of the counties according to the discriminant analysis

County	**Productivity**			
	Very low productivity	**Low productivity**	**High productivity**	**Very high productivity**
	Botoşani Neamț Suceava Vaslui Buzău Vrancea Călăraşi Dâmbovița Giurgiu Ialomița Teleorman Bistrița-Năsăud Harghita	Iaşi Brăila Tulcea Mehedinți Olt Vâlcea Arad Caraş-Severin Bihor Maramureş Satu-Mare Sălaj Alba Covasna Sibiu	Bacău Argeş Prahova Hunedoara Mureş	Constanța Galați Dolj Timiş Gorj Cluj Braşov Ilfov
Total	13	15	5	8

The hypothesis of our study according to which the order of the groups of counties in terms of productivity corresponds to the order of groups according to DA in confirmed. According to our study (see Table 7), 65.9% of the counties are correctly classified by the discriminant function. Consequently, we consider that the evaluation of the stage of development by DA is more adequate than the evaluation by the labour productivity, which, although a synthetic indicator, does not express the relationships among the determining variables of the degree of social-economic development.

4 Conclusions

The discriminant analysis gives the possibility of calculating the importance scores of the variables which influence the development level of the Romanian counties and, by this, helps in identifying the groups.

Using the discriminant analysis we have managed to identify those variables which have a strong relationship with the development level of the Romanian counties. The selected variables significantly contribute to the differentiation of the groups, namely: four variables have Sig. values lower than 0.05 (wage2003, RD2003, urb2003, phys2003) and one variable with a Sig. below 0.1 (stud2003).

The discriminant score can better evaluate the degree of the counties' development than a synthetic indicator or even a system of indicators, which lay the stress upon the level attained in the development of a phenomenon rather than on the relations, inter-conditioning which support the development of the phenomenon. Thus the discriminant analysis reproduces more accurately the actual configuration from the point of view of the degree of development.

Since it considers both the defining variables of development and the relationships among variables, the evaluation of the degree of the county development by the discriminant analysis can provide more appropriate information to help the decision-taking authorities in working out their strategy or policy of development.

References

Fisher, R. A., *The use of multiple measurements in taxonomic problems*, Annals of Eugenics No.7, 1936
Jaba, E., *Statistica*, Third Edition, Editura Economica, Bucureşti, 2002.
Jaba, E., Grama, A., *Analiza statistică cu SPSS sub Windows*, Polirom, Iaşi, 2004.
Klecka, W. R., *Discriminant Analysis*, Quantitative Applications in the Social Sciences Series, No. 19. Thousand Oaks, CA: Sage Publications, 1980.
Lachenbruch, P. A., *Discriminant Analysis*, Hafner, NY, 1975.
Saporta, G., *L'analyse discriminante*, 2007 la http://cedric.cnam.fr/~saporta/discriminante.pdf, accesat pe 20 iulie 2007.
*** *Indicators of Economic Development*, la www.bized.ac.uk/educators/16-/economics/development/presentation /indicators.ppt, accesat pe 10 martie 2007.
*** *National Human Development Report – Romania, 2007* la http://hdr.undp.org/docs/ reports/national/ROM_Romania/ ROMANIA_2007_en.pdf, accesat pe 10 septembrie 2007
*** *The American Heritage Dictionary of the English Language*, Fourth Edition, Houghton Mifflin Company, 2000, la http://education.yahoo.com/reference/dictionary/entry/ discriminant%20function, accesat pe 5 ianuarie 2007