METHODOLOGICAL BASES OF THE ANALYSIS OF THE ROMANIAN EDUCATIONAL SYSTEM

Ani MATEI[1], Silvia Elena CRISTACHE[2], Florica Georgeta ROTARU[3]
[1] PhD., Professor, National School of Political Science and Public Administration,
email: amatei@snspa.ro
[2] PhD., Professor, The Bucharest University of Economic Studies,
email: csilvia2005@yahoo.com
[3] PhD. Student, National School of Political Science and Public Administration,
email: flory_georgeta_rotaru@yahoo.com

Abstract
The analysis of social-economic phenomena that characterizes the education system in Romania must take into consideration a multitude of factors, primary and secondary, essential and nonessential, quantifiable factors and unquantifiable or approximately quantifiable, found in a relationship of mutual interdependence. We aimed to identify the main methods and analysis of the educational system in order to study the concrete manifestation of the correlations and the intensity with which they occur. Simultaneously, we presents ways of checking the presence and significance of outcome indicators in order to make sure future decisions and measures regarding the field are sustained.

Keywords
statistical analysis, One-Way ANOVA method, Two-Way ANOVA method, deviation, correlation report

Introduction

Achieving Romania's objectives of the Europe 2020 strategy is heavily dependent on the state and evolution of the education system, which in 2020 will have to be allocated 3.0% of GDP. It should be noted that this is a high goal, as far the education system being allocated only 0.49% of GDP (2013).

Reaching specific objectives question the use of a thorough analysis of socio-economic phenomena characteristic for the Romanian educational system using indicators of official statistical publications (Box 1).

In general, primary processing is done by: graphic method which points out the evolution of macroeconomic indicators of the system, the method of structural changes, the method of dynamic changes. The latter are simple processing methods either as absolute or relative indicators (intensity, structure and dynamics) or synthetic indicators (as an average).

Box 1. Main indicators of the Romanian education system

The main indicators characterizing the Romanian education system refers to material and technical base, to teachers, school children and the number of graduates. All these indicators are organized by area of residence, development regions and counties, ownership, etc.

All these indicators characterizing the Romanian system are collected and processed in accordance with the requirements of the European Statistical System EUROSTAT.

In the below section it is described,briefly, the main statistical indicators characterizing the Romanian education system, as follows:

1. Material and technical base residence areas, development regions and counties
   1.1. The number of school units by level of education;

1.2. Classrooms and school offices by level of education;
1.3. Gyms;
1.4. School Workshops;
1.5. Sports fields;
1.6. Swimming pools;
1.7. Number of PCs.
2. Teaching staff by level of education, ownership, development regions and counties
3. School population by level of education, residence areas, development regions and counties
4. Graduates by level of education, type of education, development regions and counties.
These indicators can be found in the Statistical Yearbook of Romania.

Regarding complex analysis, these indicators are the basis for determination of derived indicators, such as GDP share for education; Gross fixed capital formation of the education; The number of unemployed per 100 people with schooling, etc. We note that such indicators derivatives shall be determined according to the specific analyzes that are to be achieved.

Source: http://www.insse.ro/cms/files/publicatii/Statistica%20teritoriala/Educatie_ind_JudLoc.htm

For a more complex analysis of the connections and interdependencies between macroeconomic indicators of the Romanian educational system, elemental methods(primary processing) of analysis are often inadequate, thus requiring more complex approaches.
In order to analyse the links between factors that influence the evolution and structural changes in the educational system, firstly it is required to establish the influencing factors and then to rank them. It is followed by form-analysis that manifest causal relations between them, but also statistically measuring the intensity of the links.
After identifying the factors that influence the development of the education system we can design and build decisions and practical measures to ensure optimal conditions for its development, to enhance the influence of positive factors and to eliminate or mitigate the influence of less favourable factors.
To characterize the relationship between phenomena manifested in the functioning of the Romanian education system it is proposed using simple methods for the characterization of connections. They are easy to apply and are based on qualitative analysis of correlated variables, giving decision makers the information on the nature and essential features of the phenomena studied.
The choice of these methods for the analysis of the education system relies on the following objectives: accurate quantification of the influence of main factors; choosing from a given set the key factors, measuring the strength and direction of existing links between phenomena that characterize the Romanian education system.
To analyze the educational system in Romania we suggest: Using variation single and multifactor analysis (One-Way ANOVA and Two-Way ANOVA) for indicators used to characterize the system; significance testing of indicators used.

2. Analytical methods for measuring the links between indicators that characterize the education system

This analysis is based on studying the correlations established between specific statistical indicators provided by national statistical system mentioned above

Using simple regression to study the correlation between indicators that characterize the educational system trends consists of determining regression function parameters, achieved by applying least squares. Analytical methods take into account the real values of correlated variables respectively the specific static indicators.

Also called variation analysis, ANOVA method is underlying econometric characterization of the complexity of interdependencies of the educational system in order to determine the degree of influence of various factors or causes.

2.1 One-Way ANOVA Method

This method allows analysis of indicators characterizing the activity and educational quality depending on a single factor and implicitly comparing typical values in order to establish significant differences between them.

Using simple regression method to study the correlation between indicators characterizing the Romanian educational system requires and estimation of the regression function parameters, achieved by linear modelling described by the equation:

$$y_{x_i} = a + bx_i \tag{1}$$

The coefficients of the regression model are determined from the system of equations:

$$\begin{cases} an + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 = \sum x_i.y_i \end{cases} \tag{2}$$

The parameters "a" and "b" establish through the system of normal equations obtained by the least squared method, which is based on minimizing the squares of the deviations of the individual values recorded and the theoretical values (corresponding to the position).

$$f = \sum (y_i - a - bx_i)^2 = \text{minim} \tag{3}$$

The intensity correlation is analysed using the *indicator of correlation report*, which can be determined according to the following calculation relation:

$$R_{y/x} = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} \tag{4}$$

The correlation report (Ry/x) is a synthetic indicator used to measure the intensity of ties between both variables, but also to validate the regression models used.

2.2 Two-Way ANOVA Method

The education system is characterized by a series of complex phenomena, acting in the same or in different directions. In this case we can use multifactor regression models based on linear and non-linear functions. In this case the regression analysis follows the next steps:

- regression model development
- estimating model parameters
- checking veracity of results obtained
- the validity of the regression model obtained

In developing multifactor Two-Way ANOVA linear regression model we assume the dependency of the variables "effect" and "cause" X1, X2,... Xi,....Xn and we take into account the existence of a mutual independence of the latter.
In case factorial variables are interdependent, then multi-co linearity phenomenon occurs. According to Ragnar Frisch in his work ***Statistical Confluence Analysis by Means of Complete Regression System*** of 1934, Oslo, this represents the linear and non-linear relationship between two factorial variables, considered independent, of a correlation model. Also multi-co linearity causes the distortion of the model's parameters and its testing can be done using Farrar test and Glauber test. The first test *Farrar* checks the null-hypothesis, which states that there is no multi-co linearity in the given correlation model. The *Glauber* test allows identification of variables that are worst affected by multi-co linearity. Within multiple connections, factorial variables have different influences on outcome variables, in this context some having an important action on the phenomenon exerting effect (dependent variable), and should be taken into correlation calculations, while others have an action less important and can be neglected.
Two-Way ANOVA method allows the analysis of indicators that characterize the educational system in Romania due to several factors and comparing typical values in order to determine significant differences between them. Multiple linear regression function has the following general form:

$$Y_{x_1,x_{2,\dots}x_i,\dots.x_n} = a_0 + a_1x_1 + a_2x_2 + \dots\dots + a_nx_n + \varepsilon \tag{5}$$

in which the parameters a_0, a_1,a_2,....,a_i.....,a_n interpret similarly with the parameters of one-way linear model, they are determined using the least squared method. In this case the significance of parameters:
a_{0-} free term, with medium sized character expressing the influence of unregistered factors with constant action, except variables' influence X_1,X_2,X_3.....X_i.....X_n included in the regression model $a_{0,}$ $a_{1,}$ $a_{2,\dots,}a_{i\dots,}$ a_{n-} are regression coefficients which reflect the average change in variable Y when factorial variable X_0,X_1,X_2...X_i....X_n changes by one unit. The parameters of the regression equation are determined from the system of equations:

$$\begin{cases} a_0n + a_1\sum x_1 + a_2\sum x_2 + \dots + a_n\sum x_n = \sum y_i \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 + \dots + a_n\sum x_1x_n = \sum x_1y_i \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_0\sum x_n + a_1\sum x_1x_n + a_2\sum x_2x_n + \dots + a_n\sum x_n^2 = \sum x_ny_i \end{cases} \tag{6}$$

Specifying, defining and identifying of multi factorial linear model can be illustrated on the graph according to Figure 1.

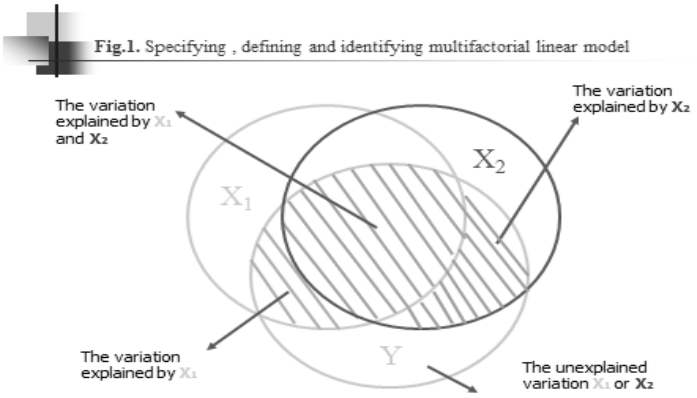

Fig. 1 Specifying, defining and identifying multifactorial linear model

The multi factorial model's estimated parameters are analysed in terms of sign, magnitude and meaning. If we consider the statistical criteria, they refer to the degree of significance of estimated parameters, the proportion of variance explained - the number of variables involved - in the total alteration of the variable effect, the existing connection between factorial variables called multi-co linearity, size of the standard error of the estimate regression parameters, autocorrelation of residual values.

For measuring concomitant influence of factorial variables on the report variables it is used *multiples correlation report* or *multiple correlation coefficients*. The multiple correlation report is noted: $R_{y/x_1,x_2....x_n}$. It measures the intensity of its implementation in multi factorial correlation models. The multiple correlation report is based on the simple correlation report and is calculated as follows: (7).

$$R_{Y/x_1,x_2\ldots x_n} = \sqrt{1-\frac{\sum(y_i - Y_{x_1,x_2,\ldots,x_i,\ldots,x_n})^2}{\sum(y_i-\bar{y})^2}} \tag{7}$$

The significance (validity) of the multi factorial econometric model is checked using the dispersion analysis or variance method (ANOVA) and the test Fisher-Snedecor(F).

In assessing the validity of the analysis model of the educational system it is checked if the variation of the cause factor/s is a good predictor for the variation of the effect factor. To measure quality adjustment for statistical regression we use two alternative indicators (standard deviation or standard error, coefficient of determination), as follows:

- **Standard deviation** (standard error) - it represent an absolute measure of quality adjustment based on regression in the sample and it is determined by the formula (8).

$$s_e = \sqrt{\frac{\Delta_e^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}} \tag{8}$$

- **Coefficient of determination** (relative indicator) determined according to formula (9).

$$R^2 = \frac{\Delta^2_{y/x}}{\Delta^2_y} = 1 - \frac{\Delta^2_e}{\Delta^2_y} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \quad (9)$$

The methods of statistical analysis presented above serve to highlight the distinctive features of the phenomena studied based on the information provided by the statistical indicators of the education system.

3. Testing (checking) the significance of the indicators used in the analysis of dependencies between them

In correlation models proposed to be used for analysis of the education system in Romania, a major problem is testing the results that were obtained, which can refer both to entirely regression function and its parameters.

(i) To test the significance of "simple correlation ratio" it is used the "Fisher" test by analysis of variance according to the formula 10:

$$F_{calc} = \frac{\sum(y_{x_i} - \bar{y})^2}{r-1} : \frac{\sum(y_i - y_{x_i})^2}{n-r} \quad (10)$$

The calculated value of the "Fisher" test compares with the critical (table) one which was determined using a probability function and then the values obtained are interpreted as follows :

If $F_{calc}>F_{tab}$ – in this case it is appreciated that the correlation ratio is a significant indicator which quantifies a real link between the cause and effect factors.

(ii) When testing the significance of multiple correlation ratio (R/ x1, x2,.....,xn) it is used the "F" test determined according to the formula 11:

$$F_{calc} = \frac{n-p-1}{p} \cdot \frac{R^2_{Y/x_1,x_2,...,x_n}}{1 - R^2_{Y/x_1,x_2,...,x_n}} \quad (11)$$

The calculated value of the "Fisher" test for multiple correlation ratio compares the critical (table) value using a probability function, and then proceed to the interpretation of the values obtained as follows:

If $F_{calc}>F_{tab}$ then it is appreciated that the multiple correlation ratio is a significant indicator and the variables included in the model are correctly identified.

Conclusions

Correlation methods proposed for the analysis of the education system in our country have the effect of quantifying the main causal factors expressed by the statistical indicators characterizing socio-economic phenomena or processes. The statistical methods and techniques to quantify the influence of factors that affect the education system in our country were chosen from a variety of statistical and mathematical procedures and instruments.

We appreciated that analysis and modeling of data related to the characterization of dependencies between indicators characterizing the education system must be based on parametric correlation methods that can be achieved using existing programs of management and database analysis. However, by including significance testing indicators used and,

respectively, resulted, analysis models get more credibility which contributes to the formulation of realistic decisions on future development.

References

1. Andrei, T.; Bourbonnais, R.; (2008), *Econometrie*, Bucharest, Economica Publishing House, pg. 111 – 123.
2. Andrei, T.; Stancu, S.; Iacob, A.I., et all, (2008), Introduction to Econometrics using EViews, Bucharest, Economica Publishing House, pg. 66- 82.
3. Bartholew, D. J., (1986): *The foundations of factorial analysis,* Biometrika, 71.
4. Boudon, R., Lazarfeld, P., (1969): *L'Analyse empiriques de la causalité*, Mouton, Paris.
5. Cristache, S.-E.(2002) – Statistics and economic theory, Course Notes, Editura ASE, Bucureşti, 2002, 252 pg., ISBN 973-594-137-6.
6. Cristache, Silvia Elena, (2009) – Introduction to econometrics tourist activity, Editura ASE, Bucureşti, 2009, 159 pg., ISBN 978-606-505-127-0.
7. Cristache, Silvia Elena, (2013), Statistics and economic theory. Course notes
8. Cristache, Silvia Elena, (2013), Introduction to econometrics tourism activity (work in manuscript).
9. Cristache, Silvia-Elena; Serban, Daniela; Vuta, Mariana, (2015), *An Analysis of The Romanian High Education System Perspectives Using Quantitative Techniques*, 5th World Conference on Learning, Teaching and Educational Leadership, WCLTA 2014, în Procedia - Social and Behavioral Sciences 186 (2015) 53 – 57, http://ac.els-cdn.com/S1877042815024313/1-s2.0-S1877042815024313-main.pdf?_tid=c60304fa-8755-11e5-8c08-00000aacb360&acdnat=1447123996_f09e26d5688037d22785b541506f3b68.
10. Calot, G.,(1975)- *Cours de statistique descriptive*, Dunod, Paris.
11. Jaba, Elisabeta, (2003) -*Statistics*, editia a III-a, Editura Economică, Bucuresti.
12. Kendall, M., Stuard, A., (1967), *The Advanced Theory of Statistics:* vol. 1. *Distribution Theory*, Ed. Griffin, London.
13. Kendall, M., Stuard, A., (1968), *Inference and Relationship* vol. 2, Ed. Griffin, London.
14. Kendall, M., Stuard, A., (1969), *Design and Analisys, and Time Series*, vol. 3, Ed. Griffin, London
15. Ragnar, Frisch, (1934), *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo.
16. Serban, Daniela, (2003), Statistics for Marketing and Business Administration, Editura ASE, Bucuresti