

---

# Computational consideration for selection of social classes in Romania

Andoria IONIȚĂ

National Institute of Statistics, Romania

---

## ABSTRACT

*Middle class is a subject discussed by almost everyone, judging it in most cases from the visible living standard's point of view: having the ownership of the dwelling, a car, making trips inside country or abroad, buying good quality and expensive goods or services and so on. But, at least in the case of our country, very often there is not a quantitative measurement of middle class, due to the fact that defining correct and reliable criteria to separate this social class from the others isn't an easy task.*

*Which are the "latent" factors which ensure each person's capability to belong to the middle class? How much this affiliation depends on the individual characteristics and how much it depends on external factors like the characteristics of the society in which the persons are living in?*

*A subtle definition of the middle class has to take into consideration several aspects, some of them more easily or more difficult to measure from the quantitative point of view. We are taking about some quantitative criteria like incomes or the number of endowment goods owned by a person, which are criteria relatively easy to estimate thought statistical methods, but also about aspects like wellbeing or social prestige, variables with a strong subjective specificity, on which there is very difficult to find an accord regarding methods of measurement between different specialists.*

*This paper presents the results of an attempt to define social classes for Romania, in order to highlight the dimensions and the social importance of the middle class in our country. The elaboration of the methodology to build the social classes starts from the definition of 11 professional categories, based on the Classification of Occupation in Romania. By using the professional categories defined, which can be considered a first instrument (or a first step) for the separation of middle class from the other ones, the present paper presents a first image of the middle class in Romania, analyzing the "chances" one person has to belong to middle class, more exactly by examining the probabilities persons with different characteristics have to belong of this part of population.*

*This paper can represent an example of good practice in using R product for the analysis of social and economic phenomena in Romania.*

**Keywords:** R, middle class, logistic regression, multinomial regression

---

---

## INTRODUCTION

Trying to measure the middle class isn't an easy task. As Régis BIGOT explains in his paper *Middle class under pressure*, "It is not easy to define middle class. The points of views under this issue are divergent, especially because there are a great variety of criteria used for separating this group of units difficult to classify: the nomenclature of Professions and social categories, incomes, level of education and, why not, patrimony owned, the status regarding employment (a stable or precarious job, partial or complet working program etc.)".

In order to highlight the dimensions and characteristics of middle class in Romania, the author defined a methodology of middle class's measurement which takes into consideration several domains and criteria in order to separate this class from at least two other classes, the one of privileged people and the other of disadvantaged people.

The dimensions and criteria took into consideration are the following:

	Domain	Criteria
1	RESOURCES	Gross annual income of population
		Level of education graduated
		Patrimony owned
2	EMPLOYMENT AND PRESENCE ON LABOUR FORCE MARKET	Professional category (derived)
3	WELFARE AND LIVING STYLE	Poverty status
		Living a situation of material deprivation
4	SOCIAL PRESTIGE	Social category of previous generation (parents' generation)

After the analyze of possibility to measure middle class using the gross annual incomes, I investigate the domain of *Employment and presence on labor force market* because the characteristics of the economic activity carried-out by a person are very strong determinants of placing a person in one or another social class.

I defined a classification of employed persons 15 years old and more in 11 professional categories using the International Classification of Occupations 2008 revision. In order to group the different occupations on professional categories, the following criteria were taken into consideration:

- the manual/ non-manual character of work;
- the type of work relation – work contract or job relation;
- existence of duties to supervise other workers;
- economic sector - agriculture, primary sector, branches with high technology and other branches of national economy;
- professional status - employee or own-account worker.

Four big groups were defined including eleven of professional categories. These are presented in Annex 1.

According with this classification, middle class is formed from the following categories:

- 
1. Specialists in technical fields
  2. Specialists in services
  3. Specialists in traditional occupations
  4. Workers and laborers with high skills and vocational training
  5. Workers and laborers with medium skills and vocational training

## METHODOLOGY

This article presents the results of using a multinomial logistic regression for the study of a person's possibilities to belong to one or another professional category from the middle class.

The model of multinomial logistic regression is an extension of the model of binomial regression, in which the dependent variable has more than two answer possibilities. This model estimates one's unit's probability of inclusion in one of the categories of a dependent variable with more than two dimensions, on the base of other variables.

The model equation is the following:

$$\ln\left(\frac{p_{i(J-1)}}{1 - p_{iJ}}\right) = \beta_{(J-1)0} + \beta_{(J-1)1}x_{i1} + \beta_{(J-1)2}x_{i2} + \dots + \beta_{(J-1)k}x_{ik} \quad (2.1)$$

The chances of success can be written as:

$$\Omega = \frac{p_{ij}}{1 - p_{iJ}} = e^{\beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \dots + \beta_{jk}x_{ik}}$$

## DATA SOURCES AND MODEL'S VARIABLES

In order to analyze the chances persons belonging to the 11 professional categories have to be part of middle class, the author used a sample representing around 20% of usual resident population of Romania, measured at reference date of Population and Housing Census in 2011. The sample included around 1.7 million employed persons.

The variables used in the model were the following:

Like **dependent** variable, CATEGORIA – Professional category was used, with the values:

- 1= Employers and leaders in big economic and social units
- 2= Employers and leaders in medium and small economic and social units
- 3= Persons with public prestige
- 4= Specialists in technical fields
- 5= Specialists in services
- 6= Specialists in traditional occupations
- 7= Workers and labourers with high skills and vocational training
- 8= Workers and labourers with medium skills and vocational training

- 
- 9= Workers and labourers with low skills and vocational training
  - 10= Own account workers in subsistence agriculture
  - 11= (un-qualified) Workers and labourers without skills or vocational training

Like independent variables the following variables were included in the model:

- SEX – *Sex* - dummy variable having the values:

- 1= male
- 2= female

- GRVR – *Age group* – continuous variable, converted to a categorical one with eight groups :

- 1= 15-24 years old
- 2= 25-34 years old
- 3= 35-44 years old
- 4= 45-54 years old
- 5= 55-64 years old
- 6= 65-74 years old
- 7= 75-84 years old
- 8= 85 years old and more

- NIVI – *Level of education* – categorical variable with seven categories :

- 1= no school graduated
- 2= primary education
- 3= gymnasium
- 4= professional or apprenticeship
- 5= high-school
- 6= post high-school or technical foreman
- 7= tertiary education

- STAO - *Occupational status* – categorical variable with seven categories :

- 1= employee
- 2= own-account worker (including employer)
- 3= retired
- 4= unemployed
- 5= pupil/student
- 6= housewife
- 7= other inactive person

- ACTP – *Activity of national economy* – categorical variable with seven categories :

- 1= agriculture, silviculture and fishing
- 2= industry

- 3= construction
- 4= transports
- 5= commercial services
- 6= social services
- 7= other activities of national economy

- VENIT – *Gross annual income* (in thousand lei) – continuous variable.

- CLASA\_M – *Social class* – categorical variable with three categories:

- 1= privileged class (categories 1,2 and 3)
- 2= middle class (categories 4-8)
- 3= disadvantaged class (categories 9-11)

## **HOW THE MIDDLE CLASS LOOKS LIKE FROM THE POINT OF VIEW OF THE CRITERIA EMPLOYMENT AND PRESENCE ON THE LABOUR FORCE MARKET ?**

The structure of employed population by professional category shows a supremacy of category 9 (Workers and labourers with low skills and vocational training) who has a share of 44.4% from the total employed population. It is followed by categories 8 (Workers and labourers with medium skills and vocational training) and 11 (Un-qualified workers and labourers without skills or vocational training) who have a share of 14.7% and respectively 13.3% from the total employed population.

According to this methodology, the privileged class would include just 2.6% from total employed population.

From the total persons 15 years old and over for which the professional category could be appointed, the persons clasified as midde class' members (the categories 4-8) represents a share of 38%. The structure by gender of this group of population is slighty in favour of men, they representing 57% from the total middle class.

### **The structure of middle class and employed population, by level of education graduated, in 2011**

*Table 1*

	Population 15 years and more	Level of education graduated						
		no school graduated	primary education	gymna- sium	profes- sional or appren- dship	high- school	post high school or technical foreman	tertiary education
A	1	2	3	4	5	6	7	8
Middle class	100.0	0.1	0.8	8.1	12.9	31.0	5.6	41.6
Employed population	100.0	1.0	5.7	22.7	17.3	29.9	3.1	20.3

*Source : author's own calculation on 2011 Population and Housing Census data*

---

From the level's of education point of view, we found that of middle class's members have a higher stock of education compared with other groups of population. The structure of middle class shows higher weights for persons who graduated high levels of education, as high-school, post high school or technical foreman or tertiary education, compared with the whole employed population. The share of population with tertiary education is almost doubled inside middle class.

## **THE USE OF LOGIT MULTINOMIAL MODEL TO ANALYZE THE MIDDLE CLASS IN ROMANIA**

The scope of using the regression model is to check the validity of the variable CLASA, which was derived from the professional categories built on the Romanian occupations, by:

- testing the degree of correlation of several variables with the variable CLASA and
- estimating the associated probabilities for each dependent variable's category.

This section presents the results of analyze obtained using the multiple logistic regression model, which, based on several predictors, estimates the probability for a person to belong to one of the population groups inside middle class.

The logistic regression models are instruments of analyse having like the main goal understanding the role of some factorial variables in explaining a result or a phenomena. The problem consists in predicting the probability with which the dependent variable records one of the possible response category, the estimation of parameters of regression equation respecting the criteria of maximum likelihood. In this case, the result which we analyse is middle class, measured using a categorical variable (*clasa*).

In a preliminary phase of using the model, we choose the reference level of the result, the category we want to use like reporting value. This procedure is done with `relevel` function, using the environment of R programming and statistical analyze. In this case we choose the following levels of reference:

- *middle class* as class (CLASA\_M=2)
- category *Workers and labourers with low skills and vocational training* (CATEGORIA=9)
- age group *35-44 ani* (GRVR=3)
- level of education *Gymnasium* (NIVI=3)
- occupational status *Employee* (STAO=1)
- group of activities of national economy *Industry* (ACTP=2)

In the multiple logistical regression model the following variable were took into consideration: the class as dependent variable and age group, level of education graduated and occupational status as independent variables.

The model's results analysis is based on calculation of chances report, which compares the chances of category m of the dependent variable (privileged class, disadvantaged class) to record a success, compared with the reference category M (middle class). The chances report compares the chances of two groups of population (persons from m category compared with those from the reference category M) which have different values of the independent variable, in the conditions in which all other independent variables remain constant.

The results of the multinomial regression model are the following:

```
> summary(regresi.a)
Call:
multinom(formula = CLASA_M ~ GRVR + NIVI + STAO, data = rpl)

Coefficients:
(Intercept)      GRVR1      GRVR2      GRVR4      GRVR5      GRVR6      GRVR7      GRVR8
1  -5.635151  -0.93241867  -0.3461950  0.06979494  0.05800443  0.5530366  0.5629748  1.036298
3  1.051890   0.03581299  -0.1263492  -0.09692655  0.24008123  2.8255004  3.9055458  4.337813
      NIVI 1      NIVI 2      NIVI 4      NIVI 5      NIVI 6      NIVI 7      STAO
1  -7.7099378  -0.9083721  -0.4597437  0.8639416  0.5829079  3.683026  0.1946983
3  0.7587821   0.5292107  -0.7600193  -1.1861791  -2.4362582  -3.647894  0.4545923

Std. Errors:
(Intercept)      GRVR1      GRVR2      GRVR4      GRVR5      GRVR6      GRVR7      GRVR8
1  0.068693630  0.027101267  0.012316639  0.01464270  0.019483521  0.08277841  0.27631663  1.0892181
3  0.007067679  0.007100387  0.005068905  0.00533855  0.007260716  0.03307307  0.09463812  0.4130574
      NIVI 1      NIVI 2      NIVI 4      NIVI 5      NIVI 6      NIVI 7      STAO
1  0.0000044936  0.34015800  0.094478988  0.071145976  0.09091596  0.067553873  0.007772566
3  0.0358768404  0.01556629  0.006347147  0.005639247  0.01076497  0.007509723  0.002956423

Residual Deviance: 1913307
AIC: 1913367
```

The exponential values of the coefficients are:

```
> exp(coef(regresi.a))
(Intercept)      GRVR1      GRVR2      GRVR4      GRVR5      GRVR6      GRVR7      GRVR8      NIVI 1
1  0.003570139  0.3936006  0.7073745  1.0722883  1.059720  1.738524  1.755888  2.818762  0.000448349
3  2.863058440  1.0364620  0.8813070  0.9076227  1.271352  16.869385  49.677189  76.539954  2.135673613
      NIVI 2      NIVI 4      NIVI 5      NIVI 6      NIVI 7      STAO
1  0.403180  0.6314455  2.3724938  1.7912395  39.76653368  1.214944
3  1.697592  0.4676574  0.3053859  0.0874876  0.02604593  1.575531
```

❖ **The chance to be in the privileged class compared with middle class**

The regression coefficient corresponding to persons from the 15–24 years age group is -0.93241867 and the exponential value of this coefficient is equal with 0.3936006. This result means that the chance a person 15-24 years old has to belong to the privileged class as to the middle one is only 39.36% from the chance a person 35-44 years old has (if both persons have the same level of education and the same occupational status). The chances increase as the age increases too, so the chance a person 65-74 years old has to belong to the privileged class and not to the middle class are 1.73 bigger compared with a person 35-44 years old.

If we analyze the chances of belonging to one or another social class from the level of education point of view, we can find that the education stock a person has is a solid premise of being placed in the privileged or middle classes: the persons who graduated the high-school, post high-school or technical foreman or the tertiary education have bigger chances to be in privileged class (and not in the middle class) compared with persons who graduated gymnasium. The chances to be in privileged class are 39 times bigger for persons having an university degree, compared with persons who graduated gymnasium. Persons without a degree have insignificant

chances to belong to the privileged class (0.045%), compared with a person who graduated gymnasium.

❖ **The risk to be in the disadvantaged class**

The risk of belonging to the disadvantaged class is much bigger for the old persons. So, this risk is 17 times bigger for persons 65-74 years old, 50 times bigger for those having 75-84 years old and 77 times bigger for persons over 85 years old, comparing with reference category (the persons 35-44 years old).

Coming to complement the idea that a higher level of education ensures an efficient “protection” towards a precarious level of life, the model’s results show that the risk of belonging to the disadvantaged class of population is bigger for persons with low level of education (without any school graduated or with primary education). A person who graduated the high-school has a risk of 30.5% from the risk of a person who graduated gymnasium to belong to the disadvantaged class of population. This risk decreases to 8.7% and 2.6% respectively, for persons who graduated post high-school or technical foreman or persons who have an university degree.

❖ **The probabilities to belong to the middle class**

Another use of the model was to estimate the probabilities associated each response category for the independent variables, by using the fitted function, in order to check the validity of the methodology of defining the social classes.

Using the first 6 records from the population sample, namely:

```
> pp <- fitted(regresia)
> head(pp)
```

	2	1	3
1	0.3427485	0.001006618	0.65624486
2	0.4187891	0.004309650	0.57690125
3	0.7752075	0.133714118	0.09107834
4	0.7752075	0.133714118	0.09107834
5	0.4422183	0.004879720	0.55290199
6	0.4187891	0.004309650	0.57690125

The model estimates the following probabilities:

```
> head(rpl)
```

	CAT	CLASA_M	SEX	VARSTA	GRVR	NIVI	OCUP	STAO	ACTP	VENIT
1	9	3	1	47	4	4	8331	1	4	14646.82
2	9	3	2	43	3	5	5151	1	5	13054.57
3	2	1	1	39	3	7	1324	1	6	21605.16
4	6	2	2	37	3	7	2631	1	5	24360.37
5	9	3	1	47	4	5	5151	1	5	11621.23
6	9	3	2	42	3	5	5151	1	5	13054.57

By analysing the individual probabilities of belonging to one of the three classes defined by the methodology, we can verify the corectness of the inclusion of several professional categories into social classes defined according to the occupations and the separation in this way of the middle class from the point of view of the characteritics of economic activity carried-out and of working place.



---

So, we found that a male person 47 years old, having a professional or apprenticeship level of education, working in transports, with an gross annual income of 14,647 lei and being classified in category 9 (Workers and labourers with low skills and vocational training) has a high probability (65.6%) to belong to the disadvantaged class of population and a small probability to belong to the middle class, which attests the inclusion of this category made in the defined class methodology as a correct one.

The fifth and sixth persons from the sample, having opposite gender, close ages, having the same occupation (5151) and the same level of education (high-school), working in the same activity of national economy and having a relatively close level of incomes, have a probability below 50% to belong to the middle class, which certify the placement of their occupation in the category 9, which is defined to be part of disadvantaged class.

The fourth person having the occupation 2631, the age of 37 years, who graduated university and who works in commercial services has a 77.5% probability to belong to the middle class, the probabilities to be part of other classes being very low (13.4% and 9.1% respectively). Again, the choice done in the defined methodology proves to be correct.

## CONCLUSION

The use of multinomial logistical regression model proved to be very useful in checking the correctness of inclusion the 434 elementary occupations included in the Classification of Occupation in Romania in the professional categories which allow to separate middle class from the perspective of the *Employment and presence on labour force market* domain. This permits to continue the study of middle class in Romania, by identifying the quantitative dimensions and the characteristics of persons who belong to middle class.

## ACKNOWLEDGMENT

This paper has been financially supported within the project entitled „**SOCERT. Knowledge society, dynamism through research**”, contract number POSDRU/159/1.5/S/132406. This project is co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013. **Investing in people!**

**The inclusion of four digits occupations in 11 professional categories  
defined by the class methodology**

Professional category	Occupation codes (4 digits)
<b>A. ELITE</b>	
1	Employers and leaders of big economic or social units 1111-1114; 1120; 1211-1213; 1219; 1222; 1223
2	Employers and leaders of medium and small economic or social units 1311-1312; 1321-1324; 1330; 1341-1346; 1349; 1411-1412; 1420; 1431; 1439
3	Persons with public authority 2120; 2612; 2641-2642; 2652; 2654-2656
<b>B. SPECIALISTS</b>	
4	Specialists in technical fields 2113-2114; 2131-2133; 2141-2146; 2149; 2151-2153; 2161-2166; 2263-2269; 2411-2413; 2421-2423; 2431-2432; 2619; 3151-3155; 3311-3315; 3321; 3331; 3352-3355; 3359; 3411-3413;
5	Specialists in services 2611; 3322-3324; 3332; 3334; 3339
6	Specialists in traditional occupations 2111-2112; 2211-2212; 2221-2222; 2230; 2240; 2250; 2261-2262; 2310; 2320; 2330; 2341-2342; 2351-2356; 2359; 2424; 2433-2434; 2511-2514; 2519; 2521-2523; 2529; 2621-2622; 2631-2636; 2643-2651; 2653; 2659; 3342-3344; 3351; 3421-3433; 4110
<b>C. WORKERS AND LABOURERS</b>	
7	Workers and labourers with high skills and vocational training 3111-3119; 3121-3123; 3131-3135; 3139; 3141-3143; 3211-3214; 3257; 3333; 3341; 3434-3435; 3511-3514; 3521-3522; 4211; 7411-7413; 7421-7422; 7511
8	Workers and labourers with medium skills and vocational training 3221-3222; 3230; 3240; 3251-3255; 3258-3259; 4120; 4131-4132; 4212-4214; 4221-4226; 4229; 4311-4313; 4321-4323; 4411; 4413-4416; 4419; 5111-5113; 5120; 5131-5132; 5141-5142; 5165; 5411-5412; 7111-7113; 7115; 7119; 7121-7124; 7126-7127; 7212-7213; 7221-7224; 7231-7233; 7311-7313; 7315; 7321-7322; 7512-7516; 7521-7523
9	Workers and labourers with low skills and vocational training 3256; 4227; 4412; 5151-5153; 5161-5164; 5169; 5221; 5223; 5230; 5241-5246; 5249; 5311-5312; 5321-5322; 5329; 5413; 5419; 6111-6114; 6121-6123; 6129-6130; 6210; 6221-6224; 7114; 7125; 7131-7133; 7211; 7214-7215; 7234; 7314; 7316-7319; 7323; 7531-7536; 7541-7544; 7549; 8111-8114; 8121-8122; 8131-8132; 8141-8143; 8151-8157; 8159; 8160; 8171-8172; 8181-8183; 8189; 8211-8212; 8219; 8311-8312; 8322; 8331-8332; 8341-8344; 8350
<b>D. PRECARIOUS CATEGORY</b>	
10	Own account workers in subsistence agriculture 6310; 6320; 6330
11	Un-qualified workers and labourers without skills or vocational training 5211-5212; 5222; 5414; 8321; 9111-9112; 9121-9123; 9129; 9211-9216; 9311-9313; 9321; 9329; 9331-9334; 9411-9412; 9510; 9520; 9611-9613; 9621-9624; 9629

---

#### Bibliography

1. Aldrich, J. H., Nelson, F. D., 1984, Linear probability, logit, and probit models, Thousand Oaks, CA: Sage.
2. Bailey, K. - Typologies and Taxonomies: An Introduction to Classification Techniques, 1994
3. Bigot, R. – Les classes moyennes sous pression, Centre de Recherche pour l'Etude et l'Observation des Conditions de Vie, cahier de recherche n. 249, Paris, 2008
4. Caragea, N., 2015, Statistică - concepte și metode de analiză a datelor, Editura Mustang, București, ISBN 978-606-652-063-8
5. Caragea, N., 2015, Alexandru A.-C., Dobre A.M., R – a Global Sensation in Data Science Romanian Statistical Review, vol. 62(2), pages 7-16, 2014
6. Dobre, A.M., Caragea, N., Alexandru A.-C., R versus Other Statistical Software, Ovidius University Annals, Economic Sciences Series, Ovidius University of Constantza, Faculty of Economic Sciences, vol. 0(1), pages 484-488, 2013
7. Fox, J., 2008, Applied Regression Analysis and Generalized Linear Models. 2nd edition. Sage, Thousand Oaks CA
8. Garson, G. D., 2011, Logistic Regression, from Statnotes: Topics in Multivariate Analysis. <http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm>.
9. Rose, D., Pevalin D., Elias P., Martin J. – Towards a European Socio-economic Classification, Final report to Eurostat of the Expert Group, UK Office for National Statistics, 2001
10. Zeileis, A., Croissant, Y., 2010, Extended Model Formulas in R: Multiple Parts and Multiple Responses, Journal of Statistical Software, 34(1), 1–13. URL <http://www.jstatsoft.org/v34/i01/>.