Determinants of Labor Force Potential in Romania

Silvia PISICĂ
National Institute of Statistics, Romania
Nicoleta CARAGEA (nicoleta.caragea@insse.ro)
National Institute of Statistics, Romania, Universitatea Ecologica

ABSTRACT

In this research study there were applied Multinomial Logistic Regression models to examine the socio-economic factors that were responsible conducting individuals to be part of the employment or not. As a result of the multinomial regression model, the most significant factor to consider here is that each one tells the effect of the predictors of risk on the probability of success in that category, in comparison to the reference category. For computing the multinomial logistic regression model it was used the multinom function from the nnet package in R.

This research will contribute to know the determinants of labor force potential in Romania. The data from a Romanian labor force survey 2013 is used for this study.

Key words: Labor Force Potential, Multinomial Regression, Labor Force Survey, R, Packages

JEL Classification: J21, C50, C87

INTRODUCTION

Taking part of the labor force is very important because not only the individual's life depends upon it but also it participates in the economic development of the country. The key issue to be discussed in this study is to analyze, through statistical tools, potential employment in Romania based on socio-economic characteristics of the population. The economic problem could be regarded as a risk analysis while an individual is employed or is part of other labor force category, as discouraged people, for example. As a result of the multinomial regression model, the most significant factor to consider here is that each one tells the effect of the predictors of risk on the probability of success in that category, in comparison to the reference category. These kind of econometric models have a different approach comparing with the parametric models, being part of the class of generalized linear model - GLM. These models have been formulated for the first time by John Nelder and Robert Wedderburn.

Multinomial logistic regression is used to predict categorical placement in or the probability of category membership on a dependent variable based on multiple independent variables. When a qualitative predictor has more than two levels, a single dummy variable cannot represent all possible values. Multinomial logistic regression is a simple extension of binary logistic regression that allows for more than two categories of the dependent or outcome variable. Like binary logistic regression, multinomial logistic regression uses maximum likelihood estimation to evaluate the probability of categorical membership. Using dummy variable approach presents no difficulties when incorporate both quantitative and qualitative predictors.

METHODS

In this research study, Multinomial Logistic Regression models were applied to examine the socio-economic factors that were responsible conducting individuals to be part of the employment or not. The explained variable of labor force potential of the respondent was considered as a categorical variable with five groups that are mutually replaceable but are not in the natural order. In Multinomial Logistic Model one category of the dependent variable would be taken as the reference category. In the present study, the "employed" was taken as the reference category and we will attempt to predict potential labour force in Romania based on a number of predictors.

Multinomial logistic regression is used to model nominal outcome variables, in which the log odds of the outcomes are modelled as a linear combination of the predictor variables.

The goal of multinomial logistic regression is to construct a model that explains the relationship between the explanatory variables and the outcome, so that the outcome of a new "experiment" can be correctly predicted for a new data point for which the explanatory variables, but not the outcome, are available. In the process, the model attempts to explain the relative effect of differing explanatory variables on the outcome.

It is assumed that we have a series of n observed data points. Let note with Y the dependent variable having J responses.

$$Y_i = \begin{cases} category-1 \\ category-2 \\ category-3 \\ \dots \\ category-J \end{cases}$$

For the observation i, the probability is:

$$p_i = \begin{cases} p_{i1} \\ p_{i2} \\ p_{i3} \\ \dots \\ p_{ij} \\ \dots \\ p_{iJ} \end{cases}$$

$$j = 1,2,\dots,J$$

When using multinomial logistic regression, one category of the dependent variable is chosen as the reference category. Separate odds ratios are determined for all independent variables for each category of the dependent variable with the exception of the reference category, which is omitted from the analysis. J – is the reference category. One simple way to arrive at the multinomial logit model is to imagine, for J possible outcomes, running J-1 independent binary logistic regression models, in which one outcome is chosen as a "pivot" and then the other J-1 outcomes are separately regressed against the pivot outcome[1].

The equation of the model is:

$$\ln\left(\frac{p_{ij}}{1-p_{iJ}}\right) = \beta_{j0} + \beta_{j1}x_{i1} + \beta_{j2}x_{i2} + \ldots + \beta_{jk}x_{ik} = \beta_j^{'}x_i$$

The odds:

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

The success probability for the observation *i*, $j < J$ is given by:

$$p_{ij} = \frac{e^{\beta_J^{'}x_i}}{1+\sum_{j=1}^{J-1} e^{\beta_j^{'}x_i}} \text{ ; când } j < J;$$

The failure probability for the observation *i* , $j = J$ is given by:

$$p_{iJ} = \frac{1}{1+\sum_{j=1}^{J-1} e^{\beta_j^{'}x_i}} \text{ ; când } j = J;$$

To interpret the results of multinomial logistic regression model, the odds ratio is calculated:

$$OR = \frac{\Omega_{(x_{ik}+1)}}{\Omega_{(x_{ik})}} = \frac{e^{\beta_{j0}+\beta_{j1}x_{i1}+\beta_{j2}x_{i2}+\ldots+\beta_{jk}(x_{ik}+1)}}{e^{\beta_{j0}+\beta_{j1}x_{i1}+\beta_{j2}x_{i2}+\ldots+\beta_{jk}x_{ik}}} = e^{\beta_{jk}}$$

The odds ratio compares the chances of two population groups characterized by different values of the independent variable (x_{ik} and $x_{ik}+1$), all other independent variables being constant. In other words, in the process, the model attempts to explain the relative effect of differing explanatory variables on the outcome.

β_{jk} - the exponential beta coefficient represents the change in the odds of the dependent variable being in a particular category relative to the reference category, associated with a one unit change of the corresponding independent variable x_{ik} .

1. For example, if the dependent variable has the category of responses (education level – low, medium, high), low level is the base category, then other two models are estimated:
medium level educated people vs. low level educated people
high level educated people vs. low level educated people

Variable selection for multinomial logistic regression is similar to those used with standard multiple regressions. The results show the logistic coefficient for each predictor variable for each alternative category of the outcome variable; alternative category meaning, not the reference category. The logistic coefficient is the expected amount of change in the logit for each one unit change in the predictor. The logit is what is being predicted. The closer a logistic coefficient is to zero, the less influence the predictor has in predicting the logit.

DATA SOURCE AND SOFTWARE USED

Data of the Labor Force Survey (LFS) for Romania conducted by the National Institute of Statistics in 2013 was used for the analysis. Data were collected over four quarters in a year to capture the effects of seasonal variations.

For computing the multinomial logistic regression model it was used the `multinom` function from the `nnet` package in R. There are other functions in other R packages capable of multinomial regression. The `multinom` package does not include p-value calculation for the regression coefficients, so we calculate p-values using Wald tests. The model summary output has a block of coefficients and a block of standard errors. Each of these blocks has one row of values corresponding to a model equation.

VARIABLES OF THE MODEL

Dependent Variable is the potential of labor force that is a categorical variable with 4 groups [npot (1=employed, 2=unemployed, 3= inactive people seeking work but not available to start work and inactive people available to start work no seeking work, 4=other inactive people – not seeking work and not available to start work)].

Categories of the dependent variable are not ordered.

Independent Variables (Predictors) are following:

- *Gender* – is a dummy variable for gender that is, Male = 1 and Female = 2

- *Age (Age Groups)* - Age variable was available in LFS 2013 as a continuous variable that was further converted into a categorical variable with different group showing six different stages of life [Age Group (1=15 to 19, 2=20 to 24, 3=25 to 34, 4=35 to 44, 5=45 to 54, 5=55 to 64)].

- *Residence area* – is a dummy variable for gender that is, Urban = 1 and Rural = 2

- *Education* - a categorical Variable of Education with 6 categories [Education (1=no formal education, 2=primary and lower secondary education, 3=upper secondary education, 4=post-secondary, non-tertiary education, 5=apprenticeships, technical or vocational education, 6=tertiary education)].

- *Ethnicity* a categorical Variable of Ethnicity with 5 categories [Ethnicity (1=Romanian, 2=Hungarian, 3=Romanic, 4=German, 5=other ethnicity)].

- *Professional status* - a categorical Variable of Professional status with 3 categories [Professional status (0=no workers, 1=salary workers, 2=other non-salary workers)].

- *Economic activity* – a categorical Variable of Economic activity with 3 categories [Economic activity (0=no economic activity, 1=agriculture, 2=industry, 3=construction, 4=commercial services, 5=social services)].

- *Occupations* - Occupation are classified into following 10 different major categories, set by International Standard Classification of Occupation ISCO 2008.

Categorical Variable of Occupations with 11 categories as follows:

No occupation	0
Managers	1
Professionals	2
Technicians and associate professionals	3
Clerical support workers	4
Service and sales workers	5
Skilled agricultural, forestry and fishery workers	6
Craft and related trades workers	7
Plant and machine operators, and assemblers	8
Elementary occupations	9
Armed forces occupations	10

- *Regions* - according to Romanian administrative territorial structure, there are 8 Regions - Categorical Variable [Regions (1=North-West, 2=Center, 3= North-East, 4=South-East, 5=South, 6=Bucharest-Ilfov, 7=South-West, 8=West)].

Note: There were eliminated from database all persons working in agriculture (economic activity=1) and having agricultural, forestry and fishery occupations (occupations = 6).

POTENTIAL LABOUR FORCE ANALYSIS BASED MULTINOMIAL LOGIT MODEL

Below we fit a multiple logistic regression model, in order to predict whether an individual will be employed, based on a number of predictors. Logistic regression models are used mostly as a data analysis and inference tool, where the goal is to understand the role of the input variables in explaining the outcome. In our case, Y is the potential labor force (*npot1*).

Before running our model, we then choose the level of our outcome that we wish to use as our baseline and specify this in the `relevel` function.

The model summary output has a block of coefficients and a block of standard errors. Each of these blocks has one row of values corresponding to a model equation.

It was used a stepwise regression model. Forward stepwise selection starts with the intercept, and then sequentially adds into the model the predictor that most improves the fit. Backward-stepwise selection starts with the full model, and sequentially deletes the predictor that has the least impact on the fit. The candidate for dropping is the variable having poor correlation with the dependent variable (in the R package[1] the `step` function uses the AIC criterion for weighing the choices, which

1. Other more traditional packages base the selection on z-score pr F-statistics, adding "significant" terms, and dropping "non-significant" terms. These are out of fashion, since they do not take proper account of the multiple testing issues. R software packages implement hybrid stepwise-selection strategies that consider both forward and backward moves at each step, and select the "best" of the two.

takes proper account of the number of parameters fit; at each step an add or drop will be performed that minimizes the AIC score).

In our case study, using **step** procedure – certain predictors were eliminated: economic activity and occupation.

The best fitted model consists of 3 equations (J-1, where J=4 categories of the dependent variable). It is compared each category of dependent variable (unemployed, inactive, others) with the reference category (employed): unemployed vs. employed, inactive vs. employed, others vs. employed;

The first equation could be written as follow:

$$
\begin{aligned}
& ln\left(\frac{P(npot1 = unemployed)}{P(npot1 = employed)}\right) = \beta_{unemployed\,0} + \\
& + \beta_{unemployed,female} \times (sex = 2) + \\
& + \beta_{unemployed,rural} \times (mediu = 3) + \\
& + \beta_{unemployed,15-19} \times (var\,sta = 2) + \beta_{unemployed,25-34} \times (var\,sta = 3) + \beta_{unemployed,35-44} \times (var\,sta = 4) + \\
& + \beta_{unemployed,45-54} \times (var\,sta = 5) + \beta_{unemployed,55-64} \times (var\,sta = 6) + \\
& + \beta_{unemployed,noEd} \times (nivs = 1) + \beta_{unemployed,lowerEd} \times (nivs = 2) + \beta_{unemployed,post\,sec\,Ed} \times (nivs = 4) + \\
& + \beta_{unemployed,aprEd} \times (nivs = 5) + \beta_{unemployed,tertiaryEd} \times (nivs = 6) + \\
& + \beta_{unemployed,Hungary} \times (nat = 2) + \beta_{unemployed,Roma} \times (nat = 3) + \beta_{unemployed,Germany} \times (nat = 4) + \beta_{unemployed,others} \times (nat = 5) \\
& + \beta_{unemployed,NW} \times (nuts = 1) + \beta_{unemployed,C} \times (nuts = 2) + \beta_{unemployed,NE} \times (nuts = 3) \\
& + \beta_{unemployed,SE} \times (nuts = 4) + \beta_{unemployed,S} \times (nuts = 5) + \beta_{unemployed,SW} \times (nuts = 7) + \beta_{unemployed,W} \times (nuts = 8)
\end{aligned}
$$

The output table of multinomial regression

```
              Df      AIC
<none>         78 26228.60
- amigo$nat    66 26241.06
- amigo$mediu  75 26336.03
- amigo$nivs   63 26595.20
- amigo$nuts   57 26703.84
- amigo$sex    75 27060.51
- amigo$varsta 63 30189.55
- amigo$stap   72 78308.30
> summary(step_regression_multinom)
Call:
multinom(formula = amigo$npot ~ amigo$sex + amigo$mediu + amigo$varsta +
    amigo$nivs + amigo$nat + amigo$stap + amigo$nuts, data = amigo)

Coefficients:
  (Intercept) amigo$sex2 amigo$mediu3 amigo$varsta1 amigo$varsta2 amigo$varsta4
2    22.29081  -2.494436    -1.858872     -7.283139    -0.4052253     -1.179418
3    20.42437  -2.031456    -1.179392     -7.372727    -0.1484405     -1.129484
4    22.27971  -1.292283    -1.619554     -4.852993     0.5905680     -1.144353
  amigo$varsta5 amigo$varsta6 amigo$nivs1 amigo$nivs2 amigo$nivs4 amigo$nivs5
2    -1.6185247     -2.142159   0.7264091  -0.9711085    2.524953   -1.273242
3    -1.3137909     -1.423490   1.8452905  -0.1128212    1.772260   -1.090083
4    -0.8539386      1.088786   2.0904372  -0.4081684    2.125546   -1.665434
  amigo$nivs6 amigo$nat2 amigo$nat3 amigo$nat4 amigo$nat5 amigo$stap1 amigo$stap2
2   1.4050234   4.124102 -1.3567383  10.101391 -0.6034218   -45.10434   -53.81055
3   0.8188141   4.337184 -0.9652566   7.998151 -0.3512376   -46.53783   -40.73485
4   0.6842043   4.275924 -1.4491315   7.213724 -0.2752149   -45.84446   -46.89199
```

```
  amigo$nuts1 amigo$nuts2 amigo$nuts3 amigo$nuts4 amigo$nuts5 amigo$nuts7 amigo$nuts8
2    1.821988   1.1034633   0.9874184    1.573522    2.537984    2.987809    0.968995
3    2.053971   1.5536400   2.1983552    1.971921    2.955187    2.956406    1.516704
4    2.290466   0.9853596   1.2370013    1.428221    2.020317    3.117641    1.591906

Std. Errors:
  (Intercept) amigo$sex2 amigo$mediu3 amigo$varsta1 amigo$varsta2 amigo$varsta4
2    13.14121   9.993522     9.988469      11.28227    0.10289773     0.6763129
3    13.14134   9.993532     9.988477      11.28230    0.10692148     0.6766890
4    13.14116   9.993494     9.988441      11.28210    0.09890281     0.6758738
  amigo$varsta5 amigo$varsta6 amigo$nivs1 amigo$nivs2 amigo$nivs4 amigo$nivs5
2     0.2897939     0.2382831   0.1548717    10.70752   0.2524612   0.6371931
3     0.2906103     0.2391014   0.1349351    10.70752   0.2795930   0.6376476
4     0.2883564     0.2343325   0.1053296    10.70748   0.2456741   0.6367364
  amigo$nivs6 amigo$nat2 amigo$nat3 amigo$nat4 amigo$nat5 amigo$stap1  amigo$stap2
2   0.7681776  0.6662212   1.727506  0.7280129  0.2359135  2.06546850 3.013564e-05
3   0.7705106  0.6666146   1.726953  0.8977346  0.2502232  0.04536687 1.517783e+00
4   0.7677311  0.6643421   1.726299  0.7596605  0.1851986  0.85390159 3.280073e-02
  amigo$nuts1 amigo$nuts2 amigo$nuts3 amigo$nuts4 amigo$nuts5 amigo$nuts7 amigo$nuts8
2   0.9810421    3.216998    3.347476    1.341817    1.232706   0.2946365   0.9860432
3   0.9831244    3.217481    3.347833    1.343010    1.233898   0.3022413   0.9879301
4   0.9799005    3.216727    3.347198    1.341258    1.232189   0.2917548   0.9846241

Residual Deviance: 26072.6
AIC: 26228.6
```

The Akaike Information Criterion (AIC) is a measure of the relative quality of a statistical model for a given set of data. That is, given a collection of models for the data, AIC estimates the quality of each model, relative to the other models. Hence, AIC provides a means for model selection. There will almost always be information lost due to using a candidate model to represent the "true" model. We wish to select, from among the candidate models, the model that minimizes the information loss. We cannot choose with certainty, but we can minimize the estimated information loss.

Table of exponentiation coefficients from the model

```
> exp(coef(step_regression_multinom))
  (Intercept) amigo$sex2 amigo$mediu3 amigo$varsta1 amigo$varsta2 amigo$varsta4
2  4794869661  0.0825430    0.1558483  0.0006870256     0.6668266     0.3074576
3   741637555  0.1311444    0.3074656  0.0006281529     0.8620513     0.3232000
4  4741944595  0.2746430    0.1979870  0.0078049790     1.8050134     0.3184300
  amigo$varsta5 amigo$varsta6 amigo$nivs1 amigo$nivs2 amigo$nivs4 amigo$nivs5
2     0.1981909     0.1174010    2.067643   0.3786631   12.490312   0.2799227
3     0.2687991     0.2408719    6.329938   0.8933103    5.884138   0.3361885
4     0.4257348     2.9706666    8.088451   0.6648669    8.377470   0.1891085
  amigo$nivs6 amigo$nat2 amigo$nat3 amigo$nat4 amigo$nat5  amigo$stap1  amigo$stap2
2    4.075622   61.81230  0.2574993  24376.901  0.5469369 2.578906e-20 4.269474e-24
3    2.267809   76.49187  0.3808855   2975.453  0.7038165 6.150006e-21 2.037420e-18
4    1.982194   71.94657  0.2347741   1357.940  0.7594089 1.230284e-20 4.315864e-21
  amigo$nuts1 amigo$nuts2 amigo$nuts3 amigo$nuts4 amigo$nuts5 amigo$nuts7 amigo$nuts8
2    6.184141    3.014588    2.684296    4.823605   12.654137    19.84216    2.635295
3    7.798806    4.728651    9.010181    7.184464   19.205315    19.22875    4.557179
4    9.879540    2.678775    3.445267    4.171274    7.540713    22.59302    4.913106
```

The Effect of Selected Variables on the Probability of Employment

Variables		Potential of labor force exp(coefficients) unemployed	discouraged people ICND+ DESC	other inactive people
Gender	Male (Reference category)			
	Female	0.0825430	0.1311444	0.2746430
Age groups	15 to 19	0.0006870	0.0006281	0.0078049
	20 to 24	0.6668266	0.8620513	1.8050134
	25 to 34 (Reference category)			
	35 to 44	0.3074576	0.3232000	0.3184300
	45 to 54	0.1981909	0.2687991	0.4257348
	55 to 64	0.1174010	0.2408719	2.9706666
Residence area	Urban (Reference category)			
	Rural	0.1558483	0.3074656	0.1979870
Education	no education	2.067643	6.329938	8.088451
	primary & lower secondary	0.378663	0.893310	0.664866
	upper secondary (Reference category)			
	post-secondary, non-tertiary	12.490312	5.884138	8.377470
	apprenticeships	0.2799227	0.3361885	0.1891085
	tertiary	4.075622	2.267809	1.982194
Ethnicity	Romanian (Reference category)			
	Hungarian	61.81230	76.49187	71.94657
	Roman	0.2574993	0.3808855	0.2347741
	German	24376.901	2975.453	1357.940
	Others	0.5469369	0.7038165	0.7594089
Region	North-West	6.184141	7.798806	9.879540
	Center	3.014588	4.728651	2.678775
	North-East	2.684296	9.010181	3.445267
	South-East	4.823605	7.184464	4.171274
	South	12.654137	19.205315	7.540713
	Bucharest-Ilfov (Reference category)			
	South-West	19.84216	19.22875	22.59302
	West	2.635295	4.557179	4.913106

Interpreting Multinomial Logit Coefficients

The intercepts give the estimated log-odds for the reference sex=male, age group=26-34, residence=urban etc.

How does the distribution of potential labour force vary as a function of covariates? To interpret the results of multinomial logistic regression model, we should calculate the odds ratio, which compares chances of category j (unemployed, discouraged people, other inactive people) for dependent variable to record a success, relative to the reference category J (employed). There are three logit equations to predict the log-odds of:

- unemployed relative to employed
- discouraged relative to employed
- other inactive relative to employed.

Interpreting the coefficient for unemployed relative to employed

Gender of the respondents played a significant role for individuals to be employed?

The estimated coefficient for the sex dummy variable in the unemployed-versus-employed equation is -2.494436.

Exponentiating, we obtain 0.0825430:

$$e^{\beta_{unemployed,female}} = e^{-2.494436} = 0.0825430 < 1$$

Thus, the relative probability of being unemployed rather than being in employment is by 91.8%[1] lower for female than for male with the same age, ethnicity, education, residential area and region. In other words, unemployment appears to be less probably for females than for males.

How can the education level explain the potential labour force in Romania?

Probability to be unemployed vs. employed for persons with primary and lower secondary education relative to persons with upper secondary education is:

$$OR = e^{\beta_{unemployed,lowerEd}} = e^{-0.9711085} = 0.378663$$

This figure can be regarded as the relative risk to be unemployed vs. employed, for the persons with lower education level, but having the same gender, age, ethnicity, or living in the same residential areas (urban or rural). On the other hand, the odds ratio for an unemployed person vs. employed is over 12 times higher for persons with post secondary education relative to upper secondary, at the same category of the other independent variables. More over, the highest level of education (tertiary) is associated to a relative risk to be unemployed vs. employed of 4 times.

The risk to be unemployed vs. employed appears to be more probable for high educated persons or, on the opposite site, for no educated at all.

A one-unit increase/decrease in the variable age group relative to the reference age group is associated with the decrease in the log odds of being in unemployed category vs. employed category of labor force. That means that being aged 20-24 not 25-34, the relative risk to be unemployed is 0.6668266. It is obviously that a youngest person vs. 25-34 years old one has the lowest risk to be unemployed (OR=0.0006870) than employed.

The relative risk for unemployed over employed is higher in any other development region in Romania than in Bucharest. The highest relative risk is in South (more than 12 times than in Bucharest) and the lowest odds ratio is calculated for West.

Interpreting the coefficient for discouraged relative to employed

Probability to be discouraged vs. employed for females relative to persons with upper males is 0.1311444. This means that females are less likely to be in the discouraged group over employed than males having the same age, ethnicity, education level and also living in the same residential area or region.

1. It is calculated as 100-8.3=91.8%

The risk of being a discouraged vs. employed is 2.3 higher for persons with tertiary education relative to upper secondary, at the same category of the other independent variables.

CONCLUSIONS AND RECOMMENDATION FOR FUTURE RESEARCH

The most important objectives of this research study were to examine the socio-economic determinants of potential labor force in Romania. This investigation was conducted using multinomial logit regression analysis, in which labour force status is modelled as a function of factors relating to socio-demographic characteristics and territorial profile issues. In multinomial logit models, where the response variable has J un-ordered categories, there are $J - 1$ parameters (or $(J-1)(k-1)$ parameters in the case of a categorical predictor with k categories) that are associated with one predictor variable. The multinomial logistic model was applied in Labor Force Survey 2013 for the analysis of potential labor force. The analysis reveals that labour force status is strongly influenced by a diverse range of factors.

The main conclusions resulted from the empirical study are the followings:

- The potential labour force could be modelled as a function of covariates which are the socio-economic characteristics of the populations' groups;
- Gender played a significant role for individuals to be employed or in other category of labour force;
- The risk to be unemployed relative to be employed appears to be more probable for high educated persons or, on the opposite site, for no educated at all;
- A youngest person vs. 25-34 years old one has the lowest risk to be unemployed than employed;
- The highest relative risk to be unemployed is in South and the lowest odds ratio is calculated for West development region.

Even the present study provides a comprehensive analysis of the determinants of labour market status among Romanian population, many aspects of well understanding the potential of labor force is not sufficiently explored in Romania. This is a continuously work for future research in order to investigate the labor force on the base of more comprehensive analysis. The present paper could serve as a benchmark for future studies including new variables in the analysis, especially those with economic impact on the potential of the labor force.

Acknowledgements

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!

REFERENCES

1. Aldrich, J. H., Nelson, F. D., 1984, Linear probability, logit, and probit models, Thousand Oaks, CA: Sage.
2. Caragea, N., 2015, Statistică - concepte şi metode de analiză a datelor, Editura Mustang, Bucuresti, ISBN 978-606-652-063-8.
3. Eeckhout, J., Jovanovic, B., 2010, Occupational Choice and Development, Journal of Economic Theory. 147(2), 657-683
4. Friedman, J., Hastie, T., Tibshirani, R., 2010, Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33 (1).
5. Garson, G. D., 2011, Logistic Regression, from Statnotes: Topics in Multivariate Analysis. http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.
6. Liebenberg, A., Hoyt, R., 2003, The determinants of enterprise risk management: evidence from the appointment of chief risk officers, Risk Management and Insurance Review. 6, 37–52
7. Lisaniler, G.F., Bhatti, F., 2005, Determinants of Female Labour Force Participation: A Study of North Cyprus. Review of Social, Economic & Business Studies. 5(6), 209 – 226
8. Pisică, S., Vasile, V., 2012, Piata muncii intre formal si informal, in Revista Romana de Economie, Vol. 34 Issue 1, pp.191.
9. 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/.

Annexes

Summary statistics for data used in multinomial logistic regression model

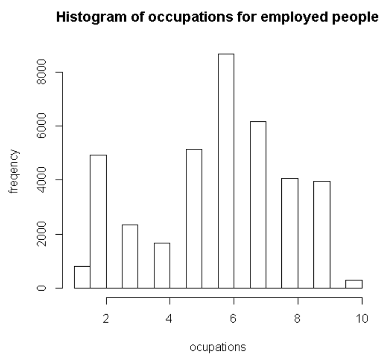
Dependent Variable	Categories	Frequencies	%
npot	1	89571	58.841
	2	6733	4.423
	3	4307	2.829
	4	51615	33.907
Explanatory Variable	Categories	Frequencies	%
sex	1	75010	49.275
	2	77216	50.725
mediu	1	84610	55.582
	3	67616	44.418
varsta	1	12172	7.996
	2	12768	8.388
	3	23715	15.579
	4	33598	22.071
	5	32311	21.226
	6	37662	24.741
nivs	1	1071	0.704
	2	42833	28.138
	3	51532	33.852
	4	5127	3.368
	5	32446	21.314
	6	19217	12.624
nat	1	140727	92.446
	2	8215	5.397
	3	2643	1.736
	4	79	0.052
	5	562	0.369
stap	0	62655	41.159
	1	61452	40.369
	2	28119	18.472
nuts	1	19898	13.071
	2	19874	13.056
	3	24890	16.351
	4	18557	12.190
	5	25156	16.525
	6	12603	8.279
	7	16961	11.142
	8	14287	9.385
act	0	62655	41.159
	1	25118	16.500
	2	20090	13.197
	3	6692	4.396
	4	23777	15.620
	5	13894	9.127
ocup	0	62655	41.159
	1	1891	1.242
	2	11618	7.632
	3	5360	3.521
	4	3814	2.505
	5	12086	7.940
	6	20472	13.448
	7	14507	9.530
	8	9817	6.449
	9	9308	6.115
	10	698	0.459

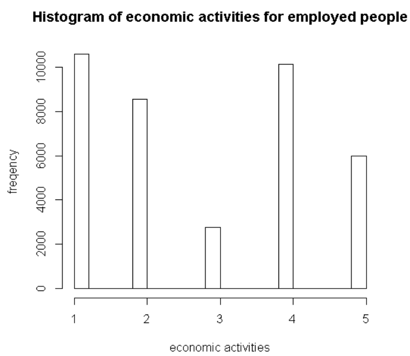
Share of female, by labor force participation (%)

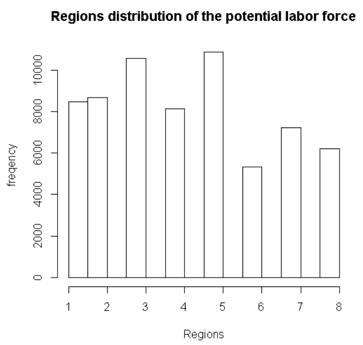
employed	51.8
unemployed	3.5
ICND+DESC	5.6
other inactive	41.8
total female	100.0

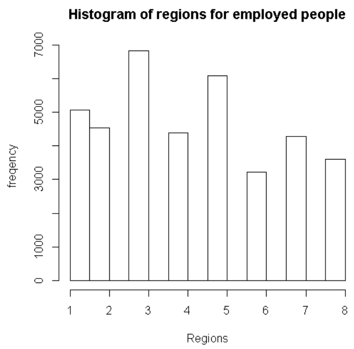
Share of male, by labor force participation (%)

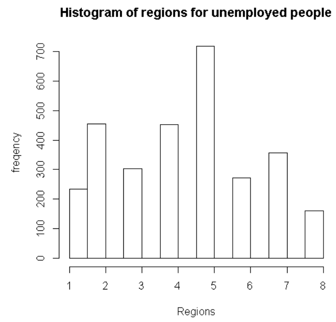
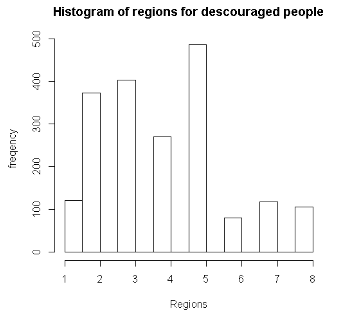
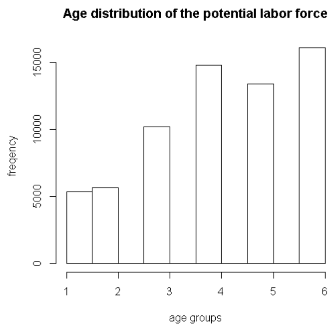
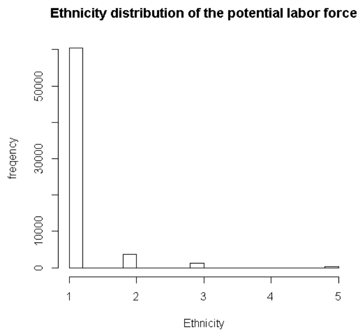
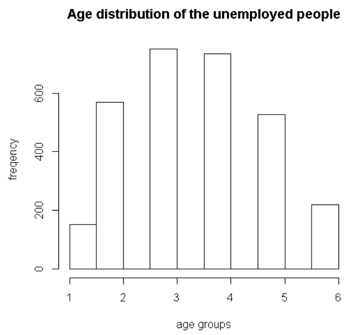
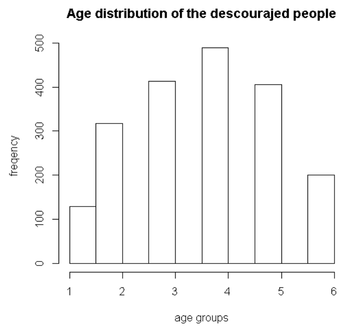
employed	66.1
unemployed	3.0
ICND+DESC	2.7
other inactive	25.8
total male	100.0

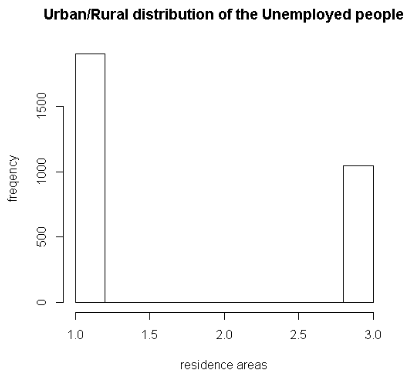


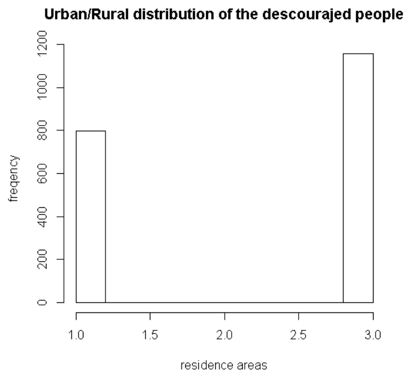


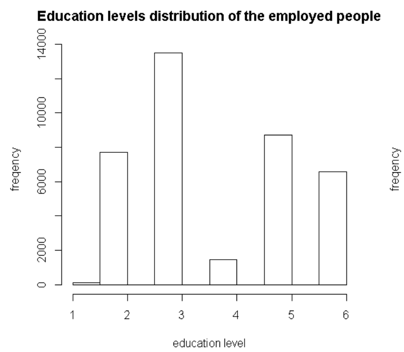


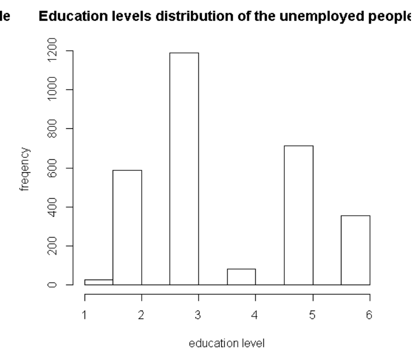


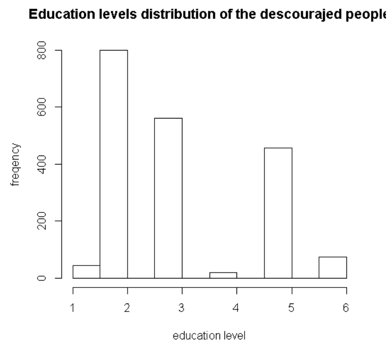

Histogram of regions for unemployed people
freqency
Regions

Histogram of regions for descouraged people
freqency
Regions

Age distribution of the potential labor force
freqency
age groups

Ethnicity distribution of the potential labor force
freqency
Ethnicity

Age distribution of the unemployed people
freqency
age groups

Age distribution of the descourajed people
freqency
age groups

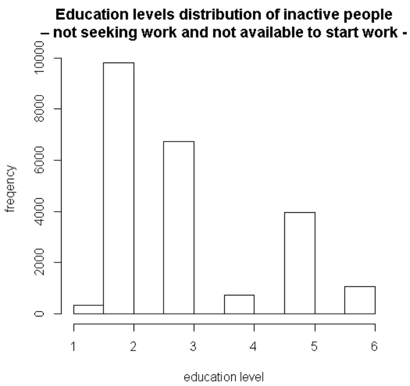

Urban/Rural distribution of the Unemployed people
freqency
residence areas


Urban/Rural distribution of the descourajed people
freqency
residence areas


Education levels distribution of the employed people
freqency
education level


Education levels distribution of the unemployed peopl
freqency
education level


Education levels distribution of the descourajed peopl
freqency
education level


Education levels distribution of inactive people
– not seeking work and not available to start work -
freqency
education level