Modelling the potential human capital on the labor market using logistic regression in R

Ana-Maria Ciuhu, Nicoleta Caragea, Ciprian Alexandru

International Conference The Use of R in Official Statistics
uRos2017
Bucharest, Romania
November 6-7, 2017

Summary

Introduction

- Analyze the potential employment in Romania based on socio-economic characteristics of the population (2 models).
- Data used is Labour Force Survey, year 2015
- The economic problem: a risk analysis while an individual is economically inactive people, being a chance to take part of the labor force.
- Binomial logit models, part of the class of GLM formulated for the first time by John Nelder and Robert Wedderburn (1972)
- The function used in R is glm().

Two models

Two types of inactives:

- economically inactive persons who are seeking for a job, but are not immediately available to start working (within 2 weeks) - Dependent variable for Model 1
- economically inactive persons who are not seeking for a job, but are immediately available to start working (within 2 weeks) - Dependent variable for Model 2
- Predictors for both models: gender, age, education level, marital status, residence area, household's structure, economic sector of the last employer and reason to decline a job offer.

Results of the Model 1 (I)

Figure: Results of the Logistic Regression Model for Inactives (Y1), reference year 2015

Covariates of the model	*Odds Ratio*	*Confidence Interval* *Lower 95%*	*Upper 95%*	*p-value*
Age (ref – gr_1)				
gr_2 (25-34 years old)	9.42e-09	NA	1.91e+199	0.9959
gr_3 (35-44 years old)	9.42e-09	NA	2.65e+171	0.9952
gr_4 (45-54 years old)	9.42e-09	NA	9.02e+172	0.9952
gr_5 (55 years or more)	9.79e-02	4.84e-03	1.41	0.0442 *
Gender (ref – male)				
female	3.34e+07	1.80e-100	NA	0.992
Residence area (ref – rural)				
urban	3.06e-01	1.51e-02	2.39e+00	0.305
Education level (ref – medium)				
low level	4.10e-07	NA	9.62e+79	0.992
high level	1.16e+00	1.39e-01	9.66e+00	0.883
Marital status (ref - single)				
married	1.35e-01	6.68e-03	1.05e+00	0.0829 .
widowed	4.95e-08	NA	2.19e+118	0.9937
divorced	4.95e-08	NA	7.41e+197	0.9961
No. of persons in the household (ref - 1)				
2 persons	1.47e+07	0.00e+00	NA	0.998
3 persons	7.93e+06	0.00e+00	NA	0.998
4 persons	5.30e+06	0.000e+00	NA	0.998
5 persons	9.99e-01	1.35e-194	7.36e+193	1.000
6 persons	9.99e-01	1.178e-200	8.48e+199	1.000
7 persons	4.28e+07	0.00e+00	NA	0.998
8 persons	9.998e-01	1.43e-256	6.96e+255	1.000
9 persons	9.99e-01	2.39e-254	4.18e+253	1.000
Economic sector (ref - industry)				
services	1.000000e+00	NA	NA	1.000
Reason (ref - distance)				
qualification	9.09	8.47	9.73	0.992
lower earnings	12.10	10.29	14.15	0.997

Source: R output on logistic regression

Results of the Model 1 (II)

- The model did not accomplish the expected empirical results;
- Perfect convergence between the dependent variables and the predictors;
- The unavailability to start work within 2 weeks is not influenced by factors included in the analysis.

Results of the Model 2 (I)

Figure: Results of the Logistic Regression Model for Inactives (Y2), reference year 2015

Covariates of the model	*Odds Ratio*	*Confidence Interval* *Lower 95%*	*Upper 95%*	*p-value*
Age (ref – gr_1)				
gr_2 (25-34 years old)	1.03	0.92	1.16	0.567
gr_3 (35-44 years old)	0.96	0.86	1.07	0.421
gr_4 (45-54 years old)	0.78	0.69	0.87	1.30e-05 ***
gr_5 (55 years or more)	1.28	1.17	1.41	5.32e-08 ***
Gender (ref – male)				
female	2.06	0.66	0.78	< 2e-16 ***
Residence area (ref – rural)				
urban	0.27	0.26	0.29	< 2e-16 ***
Education level (ref – medium)				
low level	3.06	2.87	3.26	< 2e-16 ***
high level	0.47	0.44	0.51	<2e-16 ***
Marital status (ref - single)				
married	1.50	1.07	1.24	0.000126 ***
widowed	1.96	1.80	2.14	< 2e-16 ***
divorced	0.68	0.57	0.81	1.95e-05 ***
No. of persons in the household (ref – 1 person)				
2 persons	0.77	0.63	0 94	0.01010
3 persons	0.63	0.52	0.76	2.81e-06 ***
4 persons	0.59	0.49	0.72	1.11e-07 ***
5 persons	0.63	0.52	0.77	3.76e-06 ***
6 persons	0.66	0.54	0.81	7.17e-05 ***
7 persons	0.79	0.64	1.00	0.04987 *
8 persons	0.84	0.64	1.00	0.19709
9 persons	0.69	0.53	0.91	0.00841 **
Economic sector (ref - industry)				
services	0.55	0.47	0.65	8.78e-13 ***
Reason (ref - distance)				
qualification	9.09	8.47	9.73	<2e-16 ***
lower earnings	12.10	10.29	14.15	<2e-16 ***

Source: R output on logistic regression

Results of the Model 2 (II)

- Persons are more willing to change their residence than to work on lower wages or to be re-qualified or under-qualified;
- The inactive persons are mostly older persons or people aged 25-34;
- The females have double chance more than men to be inactive seeking a job and immediately available to start working;
- The majority of inactive people live in the rural area;
- Rather people with low education and widowed or married could be inactive people;
- People which have worked before in industry sector have higher chance to be inactive than the people who have worked in the services sector.

Diagnostic of the model

Table: Results of ANOVA (chi squared) for Model 2

Covariates of the model	**ANOVA(chi squared)**
Age	<2.2e-16 ***
Gender	<2.2e-16 ***
Residence area	<2.2e-16 ***
Education	<2.2e-16 ***
Marital status	<2.2e-16 ***
No. of persons in the household	6.897e-11 ***
Economic sector	4.494e-13 ***
Reason	<2.2e-16 ***

These values indicate that every predictor improves the model.

Conclusions

- The model 2 figured out the impediments which determine inactive people not to become active on the labour market.
- The national social policies on employment should be revised. In order to attract on the labour market the inactives available to start work, employment measures should be reformulated, especially for those aged close to retirement and those in rural areas.

Thank you!

dobre.anamaria@hotmail.com
nicoleta.caragea@insse.ro
alexcipro@yahoo.com