STATISTICAL ANALYSIS OF ECONOMIC POVERTY IN POLAND USING R

Justyna Brzezińska

University of Economics in Katowice, Katowice, Poland
e-mail: justyna.brzezinska@ue.katowice.pl

Abstract: Economic poverty is one of the more common and complex problems in the modern world, as well as in Poland. This is a complex and multidimensional phenomenon, and therefore there is no single universally valid definition of poverty. This article presents a statistical analysis of economic poverty in Poland based on real data from the Central Statistical Office of Poland. An in-depth statistical analysis of the social situation of Poles will be presented, as well as an attempt to examine interdependencies in the occurrence of various forms of poverty and social exclusion in Poland. In the article, several multivariate statistical methods are presented together with the graphical presentation of results. We present a correspondence analysis with a perception map, as well as the advanced modern visualizing tool for categorical data. All the calculations were conducted using R software.

Keywords: economic poverty, multivariate statistical analysis, categorical data analysis, R software.

1. Introduction

Since the 1990s, the Central Statistical Office of Poland has regularly published data on the extent of economic poverty estimated on the basis of the results of household budget surveys, using different poverty lines. Taking into account several, not just one selected border, and the fact that none of the borders has been defined in Poland as an official border, methodologically there is no reason to prefer one of them. Each of the presented approaches has advantages and disadvantages and different interpretations. In the literature, the subject of poverty and poverty is well-known and described. The criteria for poverty have been described by Szarfenberg, Żołędowski and Theiss [2010a; 2010b], and the problem of poverty and social exclusion can be found in Lister [2007] and Damon [2012].

The rate of poverty (the extent of poverty) is the proportion of people living in households where the level of expenditure (also including the value of goods received free of charge and the value of consumption) was lower than the poverty line (www.gus.pl).

The levels of economic poverty cover three basic areas:

1) level of the subsistence minimum, adopted as the limit of extreme poverty. Minimum living is calculated by the Institute of Labor and Social Affairs (IPiSS). It only covers those needs whose satisfaction cannot be delayed and when consumption lower than that leads to biological exhaustion,

2) the statutory poverty line, defined as the amount that, under the applicable social assistance law, entitles one to apply for a cash benefit from social assistance,

3) the relative poverty line, defined as 50% of the average total household expenditure (calculated on the basis of household budget survey results).

In the case of extreme poverty and relative poverty, the original OECD equivalence scale is used for the measurement, according to which weight 1 is attributed to the first person in the household aged 14 and over; 0.7 – each next person in this age; 0.5 – each child under 14 years of age. This means that the poverty line for a four-person household of two adults and two children under the age of 14 is 2.7 times higher than for a one-person household. In the case of poverty, two thresholds should be applied in the calculation of poverty: one-person household (single-person household) and multi-person household. In the fourth quarter of 2015, the statutory poverty line for a single-person household was PLN 634 (www.stat.gov.pl).

The goal of this paper is to analyze the structure of economic poverty in Poland using multidimensional statistical analysis and categorical data analysis using R. The hierarchical methods and modern visualizing tools in the form of advanced graphs will be presented in this paper. This study will help to classify the level of economic poverty according to different criteria, and the charts will allow a detailed analysis of its structure.

2. The measurement and the extent of poverty in Poland

In this paper we present a multivariate statistical analysis of economic poverty in Poland. The empirical analysis is based on reports on poverty published by the Central Statistical Office of Poland, including studies and levels on the extent of poverty in Poland. These results are based on the Household Budget Survey. The reports provide very detailed information on the extent of poverty by: types of households, age, education of head of household, place of residence, regions, voivodships, and group of households with disabilities.

The tables below present data on the extent of poverty by socio-economic groups of households in 2012-2015 (Table 1), age (Table 2), and the level of household head education (Table 3).

Table 1 shows that the proportion of people in households living below the extreme poverty line and the relative poverty line for the years 2012-2015 has been decreasing, which means the improving situation of people at risk of poverty. This improvement is not, however, visible in the proportion of households below the

Table 1. The extent of poverty by socio-economic groups of households in 2012-2015

Specification	% of people in household with expenditure below:											
	extreme poverty boundary				relative poverty boundary				statutory poverty boundary			
Year	2012	2013	2014	2015	2012	2013	2014	2015	2012	2013	2014	2015
Total	6.8	7.4	7.4	6.5	16.3	16.2	16.2	15.5	7.2	12.8	12.2	12.2
Employees	6.2	6.4	6.5	5.6	15.3	14.9	15.2	14.4	6.9	12.0	11.9	11.7
Farmers	11.1	13.4	12.1	14.7	26.4	26.7	28.0	28.9	12.2	22.8	21.2	25.2
Own business	2.6	3.9	4.1	3.1	7.9	9.2	8.9	8.9	2.6	7.5	6.8	7.0
Retired	4.3	4.8	5.8	5.0	11.9	11.4	12.1	11.4	3.6	6.8	7.2	7.3
Pensioners	12.1	13.2	12.5	10.7	25.8	26.3	25.5	23.2	10.7	18.2	17.8	15.9
Living from other sources	22.6	21.5	21.1	17.9	41.5	39.0	36.2	36.1	24.9	34.6	29.9	30.3

Source: [http://stat.gov.pl/obszary-tematyczne/warunki-zycia/ubostwo-pomoc-spoleczna/zasieg-ubostwa ekonomicznego-w-polsce-w-2015-r-.14.3.html].

statutory poverty line for which this percentage has remained at 12.2% since 2014. These changes are also not homogeneous in the different groups in relation to the various poverty lines.

Table 2. The extent of poverty by age in 2012-2015

Specification	% of people in household with expenditure below:											
	extreme poverty boundary				relative poverty boundary				statutory poverty boundary			
Year	2012	2013	2014	2015	2012	2013	2014	2015	2012	2013	2014	2015
Total	6.8	7.4	7.4	6.5	16.3	16.2	16.2	15.5	7.2	12.8	12.2	12.2
0-17	9.8	10.1	10.3	9.0	22.0	21.8	21.8	20.6	11.9	20.3	19.3	19.3
18-64	6.4	7.0	7.0	6.2	15.5	16.5	15.5	14.8	6.3	12.0	11.0	11.0
65 and over	3.8	4.4	4.3	4.2	10.7	9.9	10.6	10.6	3.1	5.6	5.8	6.2

Source: [http://stat.gov.pl/obszary-tematyczne/warunki-zycia/ubostwo-pomoc-spoleczna/zasieg-ubo-stwa-ekonomicznego-w-polsce-w-2015-r-,14,3.html].

Table 2 shows that the proportion of people living in households below the extreme poverty line and the relative poverty line for 2012-2015 was the smallest for the proportion of households in poverty below the poverty line and the relative poverty line for people aged 18-64 in 2015, as well as the statutory poverty boundary for people over 65. The highest percentages recorded for the extreme, relative and statutory poverty line were recorded for people aged 0-17.

Table 3. The extent of poverty by the level of education of the head of household in 2012-2015

Specification	% of people in household with expenditure below:											
	extreme poverty boundary				relative poverty boundary				statutory poverty boundary			
Year	2012	2013	2014	2015	2012	2013	2014	2015	2012	2013	2014	2015
Total	6.8	7.4	7.4	6.5	16.3	16.2	16.2	15.5	7.2	12.8	12.2	12.2
Lower secondary	16.4	16.8	18.2	16.5	32.4	32.3	33.3	32.2	15.6	25.5	25.9	25.6
Basic vocational education	9.5	10.5	10.1	9.7	22.7	23.0	23.0	22.4	10.3	18.3	17.2	18.2
Secondary education	3.4	4.4	4.7	3.6	10.6	10.8	11.2	10.8	3.9	8.4	8.1	8.0
Higher	0.6	0.9	0.9	0.6	2.6	3.1	2.9	2.7	0.7	2.1	1.9	2.0

Source: [http://stat.gov.pl/obszary-tematyczne/warunki-zycia/ubostwo-pomoc-spoleczna/zasieg-ubo-
          stwa-ekonomicznego-w-polsce-w-2015-r-,14,3.html].

The analysis of Table 3 for the extent of poverty according to the level of education of the head of household, shows that the percentage of people affected by poverty is decreasing as education increases. This dependence is visible for all studied periods from 2012-2015.

3. Multivariate statistical analysis on economic poverty in Poland

This section will present several statistical methods for analyzing the qualitative data (correspondence analysis) and modern data visualization methods available in the vcd and vcdExtra of the R program. The data comes from the reports on economic poverty published by the Central Statistical Office of Poland in 2015.

This article present multi-dimensional methods of analysis of the qualitative data together with a graphical presentation of the results. The graphical presentation of the coexistence on the perception map for data describing the extent of poverty by socio-economic group of household (Table 2) uses correspondence analysis and Ward's hierarchical method is used to validate the dendrogram results.

From the analysis of the perception map and dendrogram presented in Figure 1 we can see that there are some similarities visible between categories that are analyzed. Farmers and other sources for living are close to one another, whereas employees and living from own business are in another cluster. The retired category in the perception map is near pensioners, whereas in the dendrogram it appeared in the cluster containing employees and retired. We will present an in-depth analysis using visualizing tools.

Now we present visualizing tools for data on economic poverty by socio-economic groups of households in 2015. Figure 2 consists of two plots: a sieve plot and an association plot, where we analyze the residuals and differences between the
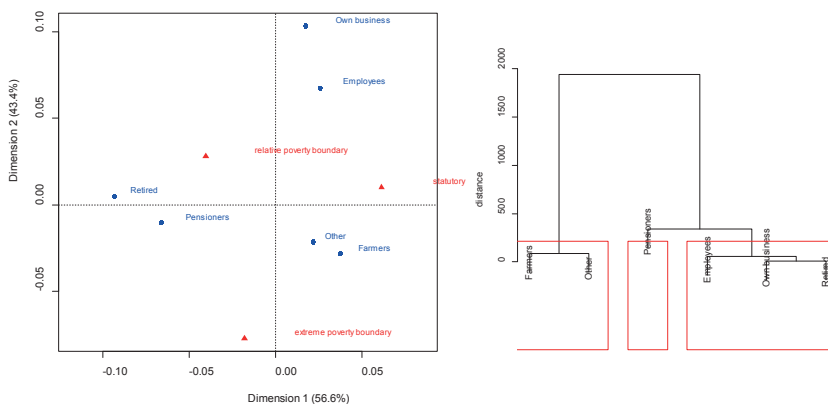
Fig. 1. Correspondence analysis and dendrogram using the Ward method for economic poverty by socio-economic groups of households
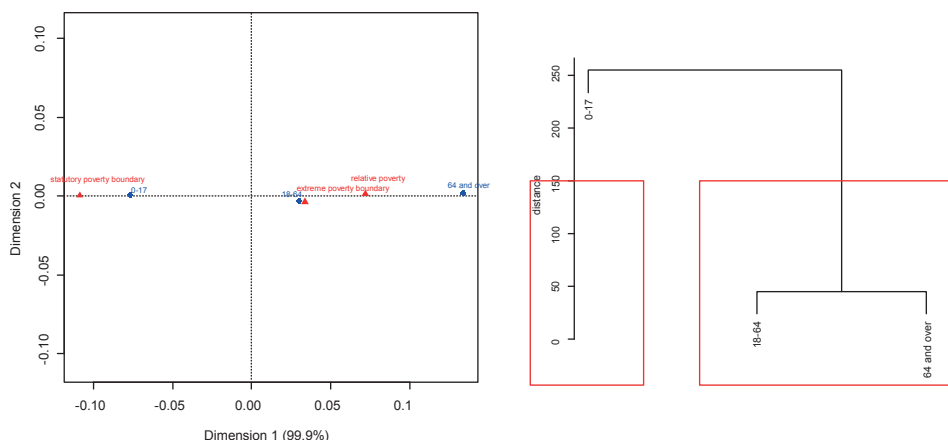
Source: own calculations.
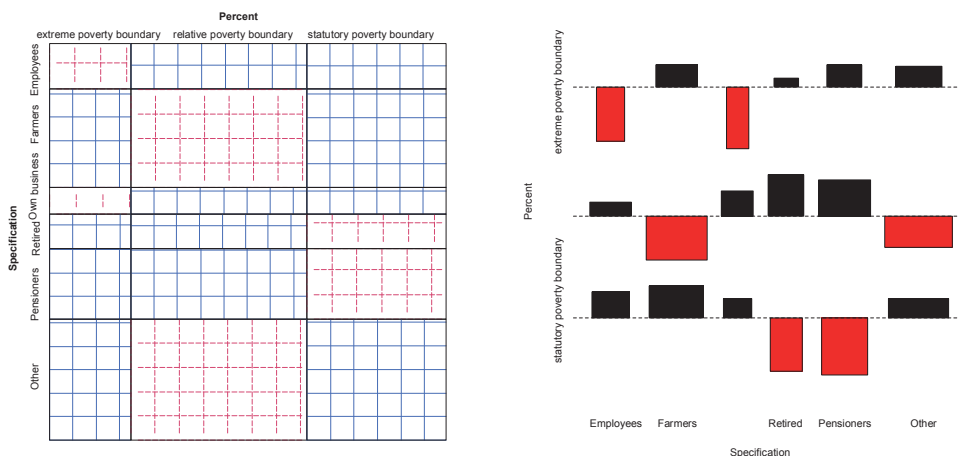
Fig. 2. Sieve and association plot for economic poverty by socio-economic groups of households

Source: own calculations.

observed and theoretical cell counts. The sieve plot with blue lines presents categories with positive differences, whereas the red lines are related to negative discrepancies. The corresponding differences for the same categories we can observe in the association plot, where the red lines are related to positive, and black – to negative differences.

Next we move to the analysis of data describing the extent of poverty by age (Table 2). The analysis of correspondence was used for the graphical presentation of

Fig. 3. Correspondence analysis and dendrogram using the Ward method for economic poverty by age
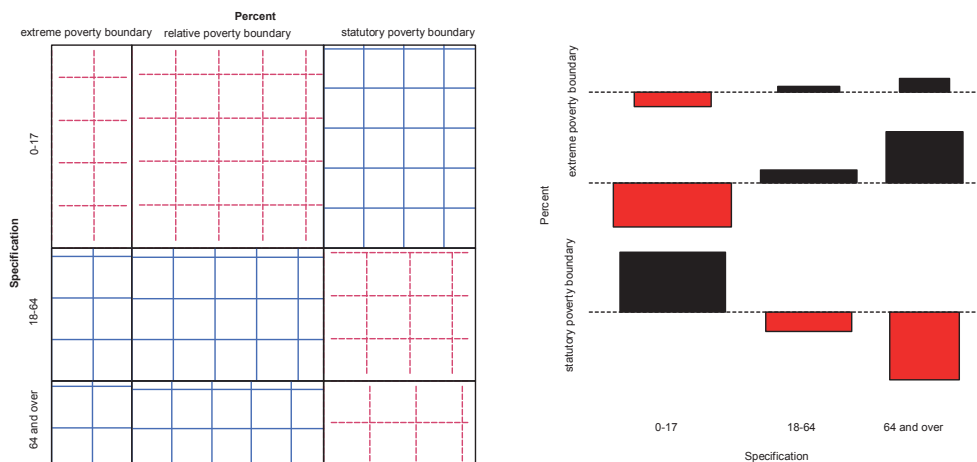
Source: own calculations.

the coexistence on the perception map and for Ward's hierarchical method was used to verify the dendrogram results (Figure 3).

From the analysis of Figure 3 (the perception map), two clusters were identified: the first cluster contains the category of the „statutory poverty boundary" and the age of „0-17", and the second is the „extreme poverty boundary „ and "18-64", as well as the „relative poverty boundary", and the age of "64 and over". These results are verified with the cluster analysis using the Ward method with two focus groups of people affected by poverty. The first cluster contains young people „0-17" and the other cluster with people aged „18-64" and „64 and over".

In the next stage of economic poverty analysis we apply modern methods of data visualization using the sieve plot and association plot. These graphs present a very clear and easy interpretation of the data structure.

From the analysis of the sieve diagram and the association plot together (Figure 4), we can see that the structure of the data set shows a similarity between the categories being examined. It is evident that empirical deviations from theoretical cell counts to those living below the „extreme poverty boundary" and the „relative poverty boundary" for people aged „0-17" are negative (red), as opposed to those living below these limits „18-64" and „65 and over". The opposite is true for people living below the „statutory poverty boundary" at the age of „0-17" for which the differences are blue for the sieve plot and black in the association.

From the cluster analysis using the Ward method we can see that the categories of poverty coverage by education (Table 4) show a disjointed focus as one cluster for "higher education" category and a second cluster for all the other categories. This

Fig. 4. Sieve and association plot for economic poverty by age
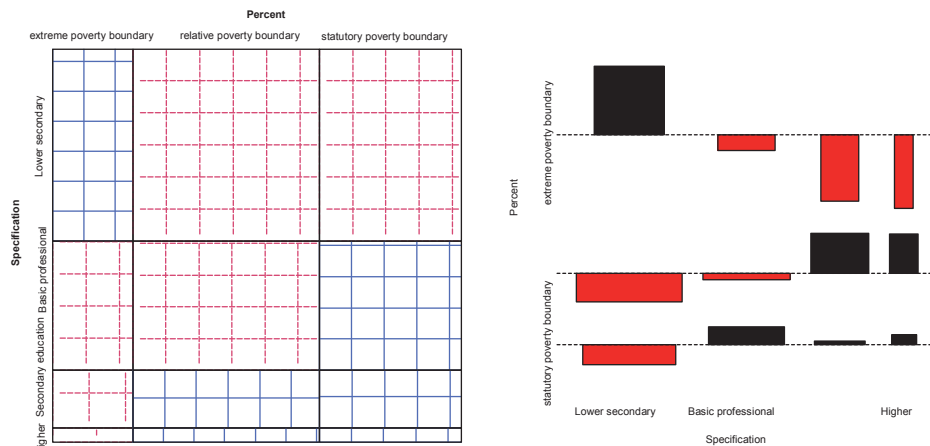
Source: own calculations.
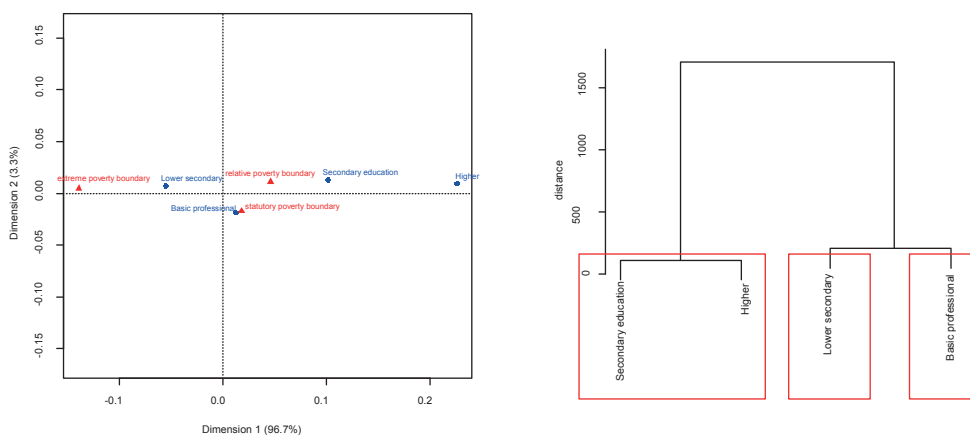
Fig. 5. Correspondence analysis and dendrogram using the Ward method for economic poverty by the level of education of the head of household

Source: own calculations.

means that people with higher education are close to category secondary education (the perception map and the dendrogram in Figure 5) and it differs from two other categories: "lower secondary" and basic vocational" education level.

From the analysis of the mosaic and association plot the similarity between residuals for higher and lower education categories (negative differences for extreme poverty and positive and statutory poverty boundary) is clearly visible. This means that these groups are most similar in terms of the deviation between empirical and

Fig. 6. Sieve and association plot for economic poverty by the level of education of the head of household

Source: own calculations.

theoretical cell counts, so that the boundary of poverty for people with higher and secondary education are among the most similar categories in the study. More differences in deviations are observed for the categories with basic vocational education level and lower secondary education.

4. Conclusions

This article presented statistical analysis of economic poverty in Poland in 2012--2015 according to different criteria in Poland based on the data from the Central Statistical Office of Poland. We applied some multivariate statistical analysis for categorical data such as correspondence analysis and verified its results with cluster analysis using the Ward method, as well as modern visualizing methods (sieve plot, association plot) that allowed for a detailed assessment of the data structure. We used R software for all calculations and graphs.

As a result of correspondence and the Ward method we can see that in the extent of economic poverty by age, there are two clusters: the first containing the category of the „statutory poverty boundary „ and the „0-17" age group, and the second is the „extreme poverty boundary „and the age"18-64", as well as „relative poverty" and "64 and over „. These results are confirmed by the screen and mosaic diagram. The analyzes of the extent of economic poverty by education have separated detached centers, where as a separate category there is higher education and the remaining categories are secondary. This means that education strongly differentiates the range of economic poverty and is an important factor influencing its level.

The methods of multidimensional statistical analysis and the presented graphs effectively illustrate the structure of the existing economic poverty in Poland. They allowed for the effective classification of different levels of economic poverty in Poland, as well as the use of visualization methods to assess its structure in detail according to several levels of poverty. In the literature of the subject, the problem of poverty, and poverty and exclusion is well-defined and well-known, however there are no deeper studies nor statistical analysis that would include the use of statistical methods to assess its level and structure and their application in R software. This article fills an existing gap in this area while also offering statistical methods in the R program and providing knowledge on poverty – a serious and crucial aspect for the economy and life in general.

Bibliography

Damon J., 2012, *Wykluczenie*, Oficyna Naukowa, Warszawa.

http://stat.gov.pl/obszary-tematyczne/warunki-zycia/ubostwo-pomoc-spoleczna/zasieg-ubostwa-eko-nomicznego-w-polsce-w-2015-r-,14,3.html (dostęp: 20.05.2017).

Lister R., 2007, *Bieda*, Wydawnictwo Sic!.

Szarfenberg R., Żołędowski C., Theiss M. (red.), 2010a, *Polityka publiczna wobec ubóstwa i wykluczenia społecznego,* Elipsa, Warszawa.

Szarfenberg R., Żołędowski C., Theiss M. (red.), 2010b, *Ubóstwo i wykluczenie społeczne – perspektywa poznawcza*, Elipsa, Warszawa.

STATYSTYCZNA ANALIZA EKONOMICZNEGO UBÓSTWA W POLSCE Z WYKORZYSTANIEM PROGRAMU R

Streszczenie: Ubóstwo ekonomiczne jest jednym z bardziej powszechnych i złożonych problemów zarówno w Polsce, jak i w świecie. Jest to problem dotykający coraz większą liczbę mieszkańców Polski. Ubóstwo jest zjawiskiem złożonym i wielowymiarowym, dlatego nie ma jego jednej powszechnie obowiązującej definicji. W niniejszym artykule przedstawiono statystyczną analizę ubóstwa ekonomicznego w Polsce. Dane poddane analizie pochodzą z raportu Głównego Urzędu Statystycznego w Polsce i dotyczą ubóstwa w ostatnich latach, opisanego wieloma zmiennymi. W opracowaniu przeprowadzono pogłębioną analizę sytuacji społecznej, a także podjęto próbę zbadania współzależności w występowaniu różnych form ubóstwa oraz wykluczenia społecznego w Polsce. W artykule wykorzystane zostaną różne metody wielowymiarowej analizy danych wraz z graficzną prezentacją wyników. Obliczenia wykonane zostaną w programie R.

Słowa kluczowe: ubóstwo ekonomiczne, wielowymiarowa analiza statystyczna, analiza danych jakościowych, program R.