Scientific Annals of Economics and Business
64 (2), 2017, 171-185
DOI: 10.1515/saeb-2017-0011

REGIONAL INCOMES STRUCTURE ANALYSIS IN SLOVAK REPUBLIC ON THE BASIS OF EU-SILC DATA

Milan TEREK*

Abstract

The paper deals with the regional incomes structure analysis in Slovak republic on the basis of European Union statistics on income and living conditions in Slovak republic data. The empirical probability mass function and empirical cumulative distribution function is constructed with aid of given sampling weights. On the basis of these functions the median, medial, standard deviation and population histogram of the whole gross household incomes for the whole Slovak republic and separately for eight Slovak regions are estimated and compared.

Keywords: regional incomes structure, sampling weights, empirical probability mass function, empirical cumulative distribution function

JEL classification: C83, R29

1. INTRODUCTION

The regional structure of incomes in Slovak republic will be analyzed on the basis of data from the European Union Statistics on Income and Living Conditions (EU-SILC) realized in Slovak republic in the year 2014. EU-SILC is an instrument aiming at collecting timely and comparable cross-sectional and longitudinal multidimensional microdata on income, poverty, social exclusion and living conditions. This instrument is anchored in the European Statistical System[1]. The start of the EU-SILC instrument was in 2004 for the EU-15. In Slovak republic EU-SILC is yearly realized from the year 2005. In general EU-SILC data are the data from complex survey.

The survey containing more components such as random sampling, stratification, clustering and so on is obviously called complex survey. A survey may be stratified with several stages of clustering and rely on ratio and regression estimation to adjust for other variables. In these cases sampling weights based on auxiliary information are commonly used to provide the correct results.

The analysis of the regional structure of incomes in Slovak republic based on using sampling weights will be studied in the paper. The using sampling weights in construction

* Department of Statistics, University of Economics in Bratislava, Slovakia; e-mail: *milan.terek1@gmail.com*.

of empirical probability mass function and empirical cumulative distribution function will be described. On the basis of these functions the estimation of population histogram, median, medial and standard deviation of the whole gross household incomes for the whole Slovak republic and separately for eight domains[2] − Slovak regions was realized.

2. MATERIAL AND METHODS

There are a lot of papers concerning the incomes distribution estimation and structure analysis, studying these problems from different regards. For example in Kloek and van Dijk (1978) the estimation of income distribution parameters is studied. The parameters of several families of distributions are estimated by means of minimum χ^2. The approach is applied on random samples taken from Dutch income-earning groups in 1973. In Ghosh *et al.* (1996) a general methodology for small domain estimation based on data from repeated surveys is studied. The results are directly applied to the estimation of median income of four-person families for the 50 states and the District of Columbia in the USA. In Sala-i-Martin (2006), the world distribution of income by integrating individual income distributions for 138 countries between 1970 and 2000 is effectuated. In Dowrick and Akmal (2005) the global income inequality is studied. In Wang and Woo (2011) the size and distribution of hidden household income in China are analyzed. In Cowell and Flachaire (2007) the statistical performance of inequality indices in the presence of extreme values in the data is analyzed. It is shown that these indices are very sensitive to the properties of the income distribution. Estimation and inference can be dramatically affected, especially when the tail of the income distribution is heavy, even when standard bootstrap methods are employed. In Atkinson and Salverda (2005) a method for using income-tax data to investigate the evolution of the highest incomes over virtually the entire 20th century is developed. In Chotikapanich *et al.* (2007) the national and regional income distributions are estimated within a general framework that relaxes the assumption of constant income within groups. A technique to estimate the parameters of a beta-2 distribution using grouped data is proposed.

Distributions of incomes or wages are obviously skewed and outliers[3] are present. Then, the interpretation power of the mean is very small[4]. Generally in such distributions the mean is not considered as appropriate measure of central tendency. Then the mean income is not convenient measure of “typical” income. The median is generally considered as good measure of central tendency in such distributions because of its stability and robustness toward outliers. Alternatively some non-traditional measures of location could be also interesting as good measures of central tendency for such distributions. The using of the trimmed mean (Piegorsch, 2015, p. 55), Winsorized mean or M-estimators is recommended[5]. Interesting results provides also traditional measures of central tendency applied on the data set from which the outliers were removed[6].

Sometimes the standard statistical methods supposing the independence and identic distribution of observations are applied to the data from complex surveys. In Lohr (2010, pp. 287-288) is stated: ”When you read the paper or book in which the authors analyze data from the complex survey, see whether they accounted for the data structure in the analysis, or whether they simply ran the raw data through non-survey statistical package procedure and reported the results. If the latter, their inferential results must be viewed with suspicion”.

2.1 Sampling weights

The sampling weights[7] allow to construct an empirical distribution for the population[8]. On the basis of this distribution the estimation of median and other quantiles[9], medial, standard deviation and population histogram is also possible. The sampling weights can be calculated using auxiliary information.

Suppose we know the size *N* of finite population *U*. Symbol *x* denotes variable under study and also its values, $U = \{1, 2, \ldots N\}$ is the set of unit indexes in the population. Symbol *S* denotes sample from the population – subset containing *n* units from *U*. Let's π_i be the probability that unit $i \in U$ will be in random sample. Sampling weights for any sampling design are defined as follows:

$$w_i = \frac{1}{\pi_i} \tag{1}$$

Sampling weight of unit *i* can be interpreted as number of units in the population represented by unit *i*.

The estimators of the parameters in cluster sampling, stratified sampling, and other sampling designs including its combinations such as for example multistage stratified sampling can be expressed by sampling weights. These weights can be modified in regard to nonresponse and coverage error[10]. Sampling weights for all observations units are equal in self-weighting surveys. Each observed unit represents the same number of unobserved units in the population. Sampling weights are not equal for all observations units in non-self-weighting surveys. If the sample is non-self-weighting, point estimates of means, totals and other quantities produced by standard statistical software will be biased. It is the case also in mentioned application. The EU-SILC sample is non-self-weighting. The capturing the structure of data is necessary in point estimation of population quantities. The using of sampling weights is needed.

2.2 Estimating an empirical probability mass function and empirical cumulative distribution function

Suppose the values for the entire population of *N* units are known. A value of probability mass function (PMF) in *x* is

$$p(x) = \frac{N_{(x)}}{N} \tag{2}$$

where $N_{(x)}$ is number of units whose value is *x*. A value of cumulative distribution function (CDF) in *x* is

$$F(x) = \sum_{y \le x} p(y) \tag{3}$$

Note that it is probability mass function and cumulative distribution function of observation from the population because the „model-free“ or „distribution-free“ approach to sample survey is under consideration[11].

Sampling weights allow to construct empirical probability mass function and empirical cumulative distribution function. Empirical probability mass function $\hat{p}(x)$ is defined by the sum of weights for all observations taking on the value *x* divided by the sum of all the weights:

$$\hat{p}(x)=\frac{\sum_{i\in S;x_i=x} w_i}{\sum_{i\in S} w_i} \tag{4}$$

Empirical cumulative distribution function $\hat{F}(x)$ is

$$\hat{F}(x)=\sum_{y\leq x}\hat{p}(y) \tag{5}$$

2.3 Plotting data from a complex survey

One from the simplest plots displaying the data distribution is histogram. If a sample is self-weighting, a regular histogram of the sample data will estimate the population probability mass function.

If a sample is non-self-weighting the sampling weights are used to construct a histogram that estimates the population histogram. The range of the data is divided into *k* classes with each class having width *b*. The height of the histogram in class *j* is

$$\text{Height }(j)=\frac{\sum_{i\in S} w_i u_i(j)}{b\sum_{i\in S} w_i} \tag{6}$$

where $u_i(j)$ = 1 if observation *i* is in class j and 0 otherwise. The denominator in formula ensures that the total area under the histogram equals 1. Such heights are obviously called the densities of relative frequencies (Wonnacott and Wonnacott, 1984, p. 107).

2.4 Estimating of some population quantities

The population parameters can be calculated on the basis of probability mass function. For example population variance is

$$\begin{aligned}\sigma_K^2 &= \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu_K)^2=\frac{1}{N}\sum_x\left(x-\sum_x xp(x)\right)^2 N_{(x)}=\sum_x\left(x-\sum_x xp(x)\right)^2 p(x)=\\ &=\sum_x x^2 p(x)-\left(\sum_x xp(x)\right)^2\end{aligned} \tag{7}$$

where μ_K is population mean.

Any population quantity can be estimated from the empirical probability mass function $\hat{p}(x)$ or from empirical cumulative distribution function $\hat{F}(x)$. For example population variance can be estimated by:

$$\hat{\sigma}_K^2 = \sum_x x^2 \hat{p}(x) - \left(\sum_x x\hat{p}(x) \right)^2 \quad (8)$$

The finite population median is defined to be value $\tilde{\mu}$ satisfying $F(\tilde{\mu}) = \frac{1}{2}$ if such a value exists. Otherwise, a population median is any value in the interval $[\tilde{\mu}_1, \tilde{\mu}_2]$, where $\tilde{\mu}_1$ is the largest value of *x* in the population with $F(x) < \frac{1}{2}$ and $\tilde{\mu}_2$ is smallest value of *x* with $F(x) > \frac{1}{2}$. In general, Q_p is 100·*p* % quantile (percentile) if $F(Q_p) = p$, if such a value exists, otherwise, $Q_p \in [a, b]$ where *a* is the largest population value of *x* with $F(x) < p$ and *b* is the smallest value of *x* with $F(x) > p$. If $p < \frac{1}{N}$, Q_p is the smallest value of *x* and if $p > 1 - \frac{1}{N}$, Q_p is the largest value of *x*.

Population quantiles are estimated as follows. Since the empirical cumulative distribution function $\hat{F}$ is a step function, the interpolation is usually needed to find a unique value for the quantile. Let y_1 be the largest value in the sample for which $\hat{F}(y_1) \le p$ and let y_2 is smallest value in the sample for which $\hat{F}(y_2) \ge p$. Then:

$$\hat{Q}_p = y_1 + \frac{p - \hat{F}(y_1)}{\hat{F}(y_2) - \hat{F}(y_1)} (y_2 - y_1) \quad (9)$$

We will formulate the relations enabling to estimate medial with aid of sampling weights. Medial (*Ml*) is such value for which the sum of variable values less or equal to *Ml* is equal to the half of variable total. It can be proven that if all values of variable are nonnegative then: $Ml \ge Q_{0,5}$ (Dagnelie, 1998, p. 81). The sum of variable values for all observations taking on the value *x* we will call the class total. The medial is calculated as median but on the basis of class totals instead of frequencies. Empirical probability mass function $\hat{p}_{Ml}(x)$ in this case can be defined as:

$$\hat{p}_{Ml}(x) = \frac{\sum_{i \in S; x_i = x} w_i x_i}{\sum_{i \in S} w_i x_i} \quad (10)$$

Empirical cumulative distribution function $\hat{F}_{Ml}(x)$ is then:

$$\hat{F}_{Ml}(x) = \sum_{y \le x} \hat{p}_{Ml}(y) \quad (11)$$

Let y_1 be the largest value in the sample for which $\hat{F}_{Ml}(y_1) \le 0{,}5$ and let y_2 is smallest value in the sample for which $\hat{F}_{Ml}(y_2) \ge 0{,}5$. Then the medial can be estimated by:

$$\hat{M}l = y_1 + \frac{0{,}5 - \hat{F}_{Ml}(y_1)}{\hat{F}_{Ml}(y_2) - \hat{F}_{Ml}(y_1)}(y_2 - y_1) \tag{12}$$

The medial provides in some application areas very interesting interpretation possibilities.

Note that estimators constructed using this method are not necessarily unbiased or numerically stable. For example the estimator $\hat{\sigma}_K^2$ of the population variance is sensitive to round off error. Despite of it, the statistics calculated using weights are much closer to the population quantities as in not weighting case (Lohr, 2010, p. 293).

3. ANALYSIS OF REGIONAL STRUCTURE OF INCOMES ON THE BASIS OF EU-SILC 2014 DATA

The analysis of regional structure of incomes was effectuated on the data from the survey EU-SILC realized in Slovak republic in 2014 (EUROSTAT, 2007). The stratified two-stage survey is used in Slovak republic. A stratification was effected with two stratification variables – region and settlement size. There are eight regions in Slovak republic. Bratislava, Trnava, Trenčín and Nitra in western Slovakia, Žilina and Banská Bystrica in central Slovakia, Košice and Prešov in eastern Slovakia. The survey EU-SILC 2014 was effectuated on the sample of 6,010 households, 5,490 households and 13,433 individuals 16 and more years old were included to database. Sampling weights were calculated and modified with respect to nonresponse. These weights can be used to inference about the population of Slovak households. Other modified sampling weights involve individuals. In general EU-SILC sample data are the data from non-self-weighting survey.

Data from EU-SILC 2014 are concentrated in many sets. Each household has one identification number. The analysis of the whole gross household incomes in eight domains – Slovak regions was realized. The mentioned regions correspond with values of one from stratification variables. Firstly the matching of needed data – sampling weights and whole gross household incomes was effected according to household numbers. Then the matched data were distributed according to regions. Eight sets of data were obtained, one for each region. Each region was analyzed separately. The values of the empirical probability mass function were calculated according to (4) and on the basis of that the values of the empirical cumulative distribution function were calculated by relation (5) for the whole Slovak republic and separately for each region.

The estimate of median whole gross household income was calculated according to relation (9) for the whole Slovak republic and separately for each region. The estimate of population median whole gross household income for the whole Slovak republic in the year 2014 equals 13,305.83 euros. The obtained results for regions are in Table no. 1. This Table contains also the ordering of regions according to the median whole gross household income.

Then the values of the empirical probability mass function $\hat{p}_{Ml}(x)$ were calculated according to (10) and on the basis of that the values of the empirical cumulative distribution function $\hat{F}_{Ml}(x)$ were calculated by relation (11) for the whole Slovak republic and separately for each region. The estimate of medial whole gross household income was calculated according to relation (12) for the whole Slovak republic and separately for each region. The estimate of medial whole gross household income in 2014 for the whole Slovak republic was 20,355.80 euros and percentage of households having incomes less or equal to medial in Slovak republic was 74.15%. The obtained results for regions are in Table no. 2. For example in the year 2014, in region Bratislava, the half of the incomes total was distributed among 76.71 % of "poorer" households (having incomes less or equal to 24,874.65 euros), the second half of incomes total was distributed among 23.29 % of "richer" households (having incomes greater or equal to 24,874.65 euros). For example in region Banská Bystrica, the half of the incomes total was distributed among 74,88 % of "poorer" households (having incomes less or equal to 18,301.86 euros), the second half of incomes total was distributed among 25.12 % of "richer" households (having incomes greater or equal to 18,301.86 euros).

Table no. 1 – Regional structure of median whole gross household income in 2014

Region number	Region name	Estimate of median whole gross household income in 2014 (Euros)	Order of region according to median whole gross household income
1	Bratislava	14,491.37	1.
2	Trnava	13,969.12	4.
3	Trenčín	14,368.47	2.
4	Nitra	12,379.67	7.
5	Žilina	14,054.85	3.
6	Banská Bystrica	11,746.41	8.
7	Prešov	13,595.22	5.
8	Košice	13,118.16	6.

Source: own

Table no. 2 – Regional structure of medial whole gross household income in 2014

Region number	Region name	Estimate of medial whole gross household income in 2014 (Euros)	Order of region according to medial whole gross household income	Percentage of households having incomes less or equal to medial
1	Bratislava	24,874.65	1.	76.71
2	Trnava	20,331.04	5.	72.44
3	Trenčín	20,555.89	2.	73.40
4	Nitra	19,364.40	6.	74.61
5	Žilina	20,492.41	3.	73.88
6	Banská Bystrica	18,301.86	7.	74.88
7	Prešov	20,489.51	4.	73.34
8	Košice	18,193.63	8.	71.90

Source: own

The dispersion of incomes in regions was characterized by standard deviation. The population variance was estimated according to (8). Then the standard deviation estimate was calculated as square root of variance estimate. The estimate of standard deviation of the whole gross household income in 2014 for the whole Slovak republic was 11,893.55 euros. The obtained results for regions are in Table no. 3.

Table no. 3 – Regional structure of standard deviation of the whole gross household income in 2014

Region number	Region name	Estimate of Standard deviation of the whole gross household income in 2014 (Euros)	Order of region according to standard deviation of the whole gross household income
1	Bratislava	17,274.84	1.
2	Trnava	10,513.97	7.
3	Trenčín	11,808.75	2.
4	Nitra	11,240.05	3.
5	Žilina	11,106.92	4.
6	Banská Bystrica	10,933.77	5.
7	Prešov	10,652.63	6.
8	Košice	8,970.23	8.

Source: own

Finally, the sampling weights were used to construct a histogram that estimates the population histogram for the whole Slovak republic and for each region separately. The widths of classes in histograms are 10,000 euros. The densities of relative frequencies in histograms were calculated according to (6). The area of rectangle in the histogram is equal to relative frequency. The number of Slovak households in 2014 can be estimated by the sum of sampling weights: $\sum_{i \in S} w_i = 1{,}850{,}842$. When the number of households is known, the number of households in the classes can be easily calculated. In Table no. 4 are densities of relative frequencies and numbers of households in defined classes.

Table no. 4 – Distribution estimating distribution of Slovak households according to the whole gross household income in 2014

Income (euros)	Density of relative frequency	Number of households
-10,000	0.0000368000	681,765
10,000-20,000	0.0000364000	673,063
20,000-30,000	0.0000178000	329,995
30,000-40,000	0.0000056500	104,565
40,000-50,000	0.0000019700	36,541
50,000-60,000	0.0000006410	11,861
60,000-70,000	0.0000002720	5,038
70,000-80,000	0.0000000747	1,381
80,000-90,000	0.0000000890	1,647
90,000-100,000	0.0000000968	1,791
100,000-	0.0000001730	3,195
Sum		**1,850,842**

Source: own

The histogram estimating population histogram for the Slovak republic is presented in Figure no. 1. As can be seen in Figure no. 1, the greatest proportion of households in Slovak republic has incomes less or equal to 10,000 euros, the households having incomes greater than 10,000 and less or equal to 20,000 euros are also frequent. The proportions of incomes greater than 70,000 euros are not discernible in the histogram.

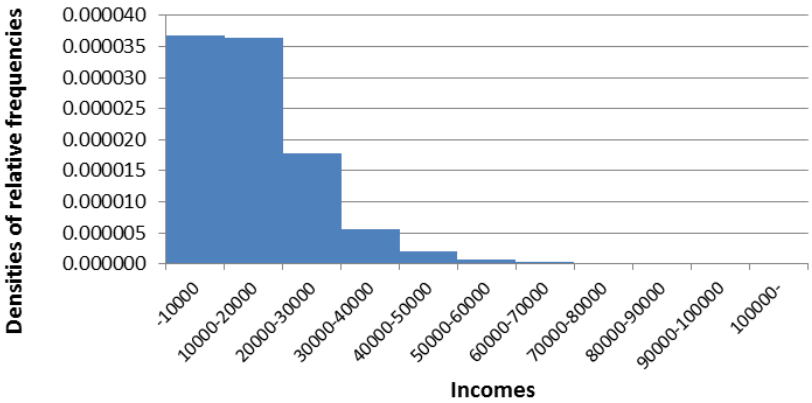

Figure no. 1 – Histogram of the Slovak households distribution according to the whole gross income in 2014

The greatest proportion of households has incomes less or equal to 10,000 euros also in Bratislava region (in Figure no. 2), but in this histogram also the households having incomes greater than 70,000 euros are in discernible proportions. The number of households in Bratislava region is estimated to be equal to 245,997.

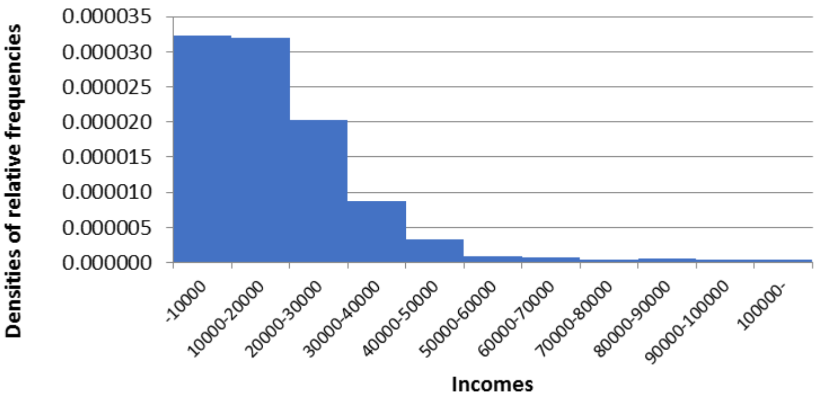

Figure no. 2 – Histogram of the Bratislava region households distribution according to the whole gross income in 2014

The greatest proportion of households in the region Trnava has incomes greater than 10,000 and less or equal to 20,000 euros (see Figure no. 3). That is difference in comparison to the whole Slovak republic and also to Bratislava region. The discernable proportions of incomes greater than 70,000 and less or equal than 80,000 and also greater than 90,000 and less or equal to 100,000 are present in that region. The number of households in Trnava region is estimated to be equal to 191,018.

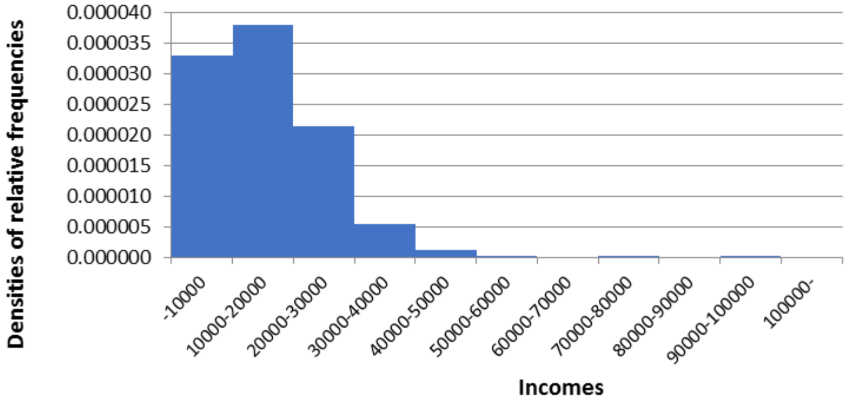

Figure no. 3 – Histogram of the Trnava region households distribution according to the whole gross income in 2014

The greatest proportion of households in the region Trenčín has incomes less or equal to 10,000 that is the most frequent are the "poorest" households but on the other hand there are also the households having incomes greater than 100,000 euros in discernible proportion (see Figure no. 4). The number of households in Trenčín region is estimated to be equal to 207,585.

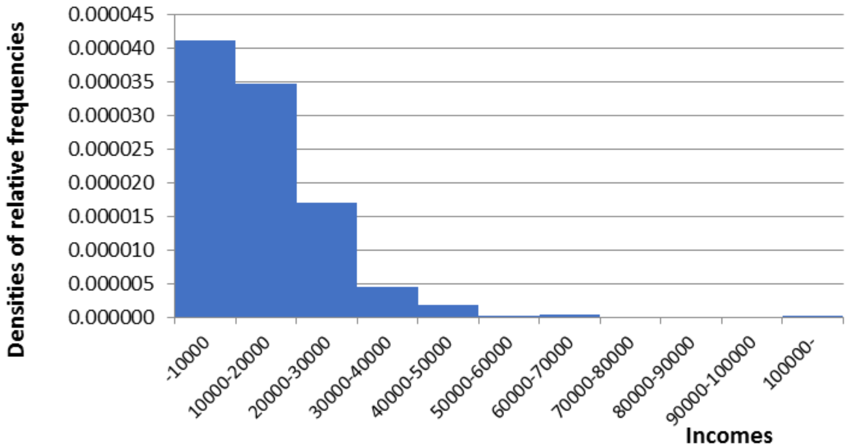

Figure no. 4 – Histogram of the Trenčín region households distribution according to the whole gross income in 2014

The greatest proportion of households in the region Nitra has incomes greater as 10,000 and less or equal to 20,000 euros (see Figure no. 5). The distribution is very similar to distribution of Trnava region, only difference is the discernible proportion of the "richest" households, having incomes greater than 100,000 euros. The number of households in Nitra region is estimated to be equal to 244,825.

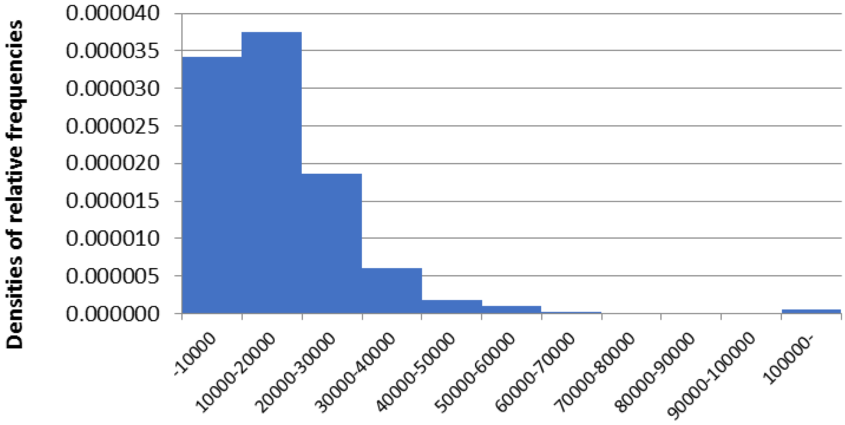

Figure no. 5 – Histogram of the Nitra region households distribution according to the whole gross income in 2014

The greatest proportion of households in the region Žilina has incomes greater as 10,000 euros and less or equal to 20,000 euros (see Figure no. 6). There is not the discernable proportion of households having incomes greater than 80,000 euros. The number of households in Žilina region is estimated to be equal to 218,788.

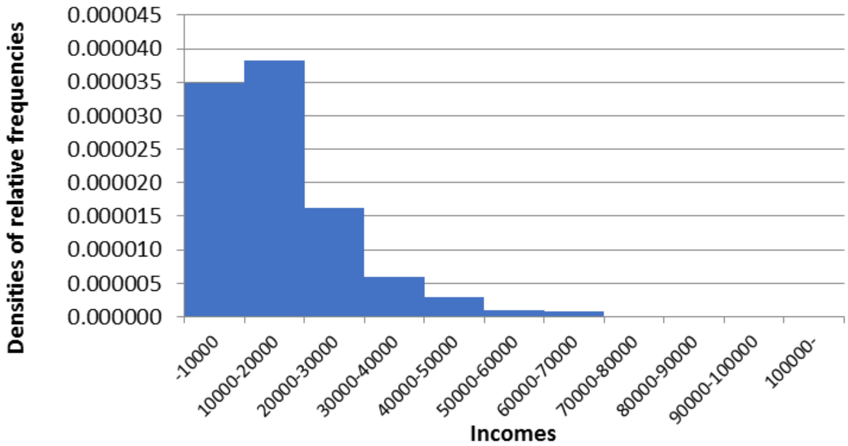

Figure no. 6 – Histogram of the Žilina region households distribution according to the whole gross income in 2014

The greatest proportion of households in the region Banská Bystrica has incomes less or equal to 10,000 euros (see Figure no. 7). There is also the discernable proportion of households having incomes more than 80,000 euros. The number of households in Banská Bystrica region is estimated to be equal to 239,708.

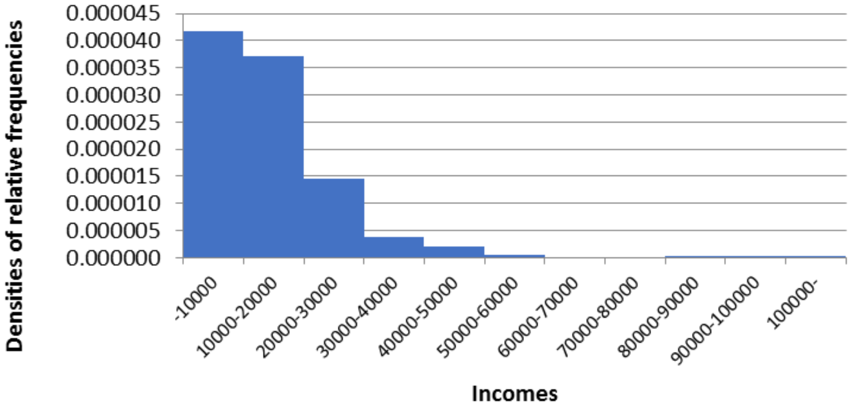

Figure no. 7 – Histogram of the Banská Bystrica region households distribution according to the whole gross income in 2014

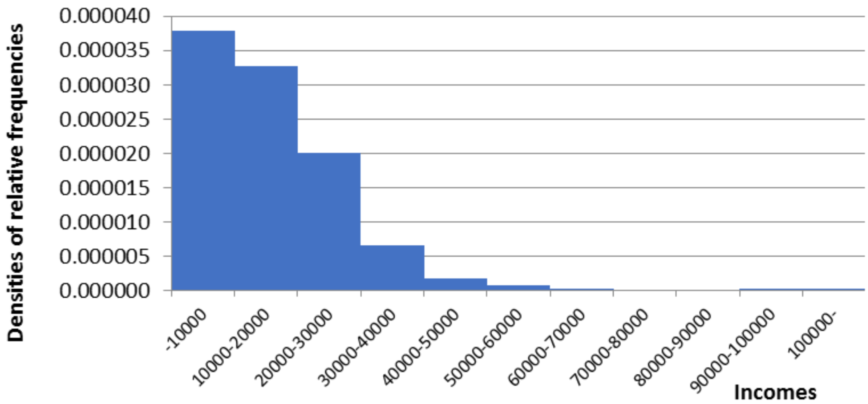

Figure no. 8 – Histogram of the Prešov region households distribution according to the whole gross income in 2014

The greatest proportion of households in the region Prešov has incomes less or equal to 10,000 euros (see Figure no. 8). There is not the discernible proportion of households having incomes greater than 60,000 euros. On the other hand the proportion of households with "middle" incomes, for example greater than 20,000 and less or equal to 30,000 is about 20%, in Banská Bystrica region only less than 15%. The number of households in Prešov region is estimated to be equal to 237,454.

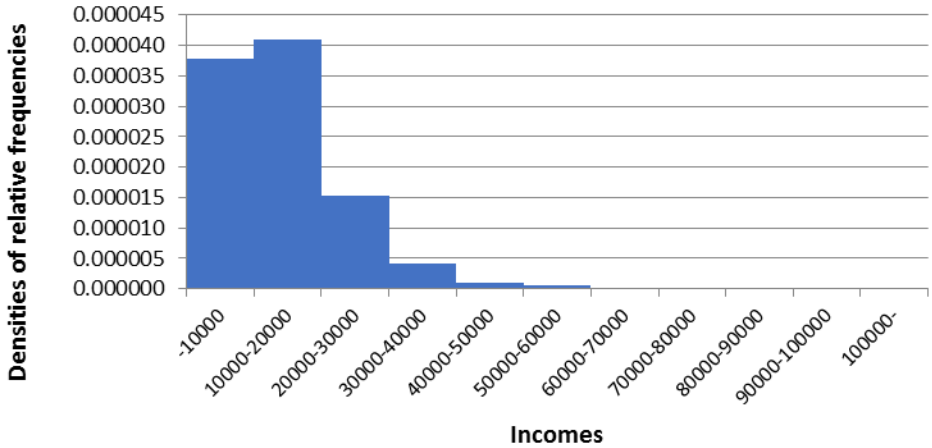

Figure no. 9 – Histogram of the Košice region households distribution according to the whole gross income in 2014

The greatest proportion of households in the region Košice has incomes greater than 10,000 euros and less or equal to 20,000 euros (see Figure no. 9). There is not the discernible proportion of households having incomes greater than 60,000 euros. The number of households in Košice region is estimated to be equal to 265,467.

4. CONCLUSIONS

All calculations were realized in Excel 2013. The obtained ordering of regions according to median whole gross household income is very interesting. Obviously the great difference among Bratislava region with the capital of Slovakia Bratislava and the rest of Slovakia is expected. The analysis results show that the difference between first Bratislava and second Trenčín regions is not very large. The median household incomes of the third Žilina, fourth Trnava and fifth Prešov are also very close. There are bigger differences among last three regions. The median whole gross household income of Banská Bystrica is surprisingly low. The regions Bratislava, Trenčín, Žilina, Trnava, Prešov have the median whole gross household income greater and the regions Košice, Nitra and Banská Bystrica less than in the whole Slovak republic.

In the analysis based on medial whole gross household income, the results of regional ordering are not very different. The first three places are occupied by the same regions as according to median, the changes in the other places of ordering are only moderate. The regions Bratislava, Trenčín, Žilina, Prešov have the medial whole gross household income greater and the regions Trnava, Košice, Nitra and Banská Bystrica less than in the whole Slovak republic. The finding that in all Slovak regions the half of the incomes total is distributed among 71.90 – 76.71% of "poorer" households and the second half of incomes total is distributed among the rest of "richer" households is very interesting. The differences among Slovak regions in this indicator are only moderate. The regions Bratislava, Nitra and Banská Bystrica have percentage of households having incomes less or equal to medial greater and the regions Trnava, Trenčín, Žilina, Prešov and Košice less than Slovak republic percentage.

The regional ordering according to standard deviation follows approximately the ordering according to median and medial. Only Bratislava region has markedly greater standard deviation of incomes – 17,274.84 euros, what is natural because this region includes the capital of Slovakia with a lot of central institutions. Others regions have not very different standard deviations of incomes.

The distribution of Slovak households according to the whole gross household income in 2014 is interesting on the level of the whole Slovak republic as well as on the level of regions. The obtained information can be very useful for example for some marketing studies.

The application of correct methodology of estimation is very important in the context of the data from complex surveys analyses. It is clear that the estimates obtained with aid of finite weights which allow the used sample design, nonresponse and potentially also coverage error better reflects the reality.

Acknowledgements

The paper was supported by grants from Grant Agency of VEGA no. 1/0092/15 entitled "Modern Approaches to Complex Statistical Surveys Design" and no. 1/0393/16 entitled „European Union in Post Crisis Period – Macro and Microeconomic Aspects".

References

Atkinson, A. B., and Salverda, W., 2005. Top Incomes in the Netherlands and the United Kingdom over the 20th Century. *Journal of the European Economic Association, 3*(4), 883-913. doi: http://dx.doi.org/10.1162/1542476054430816

Barnett, V., and Lewis, T., 1994. *Outliers in Statistical Data*. Hoboken: Wiley and Sons.

Chotikapanich, D., Griffiths, W. E., and Rao, D. S. P., 2007. Estimating and Combining National Income Distributions Using Limited Data. *Journal of Business & Economic Statistics, 25*(1), 97-109.

Cochran, W. G., 1977. *Sampling Techniques*. New York: J. Wiley and Sons.

Cowell, F. A., and Flachaire, E., 2007. Income distribution and inequality measurement: The problem of extreme values. *Journal of Econometrics, 141*(2), 1044-1072. doi: http://dx.doi.org/10.1016/j.jeconom.2007.01.001

Dagnelie, P., 1998. *Statistique Theorique et Appliquee. Tom 1 - Statistique Descriptive et Bases de l' Inference Statistique*. Paris: DeBoeck and Larcier.

Dowrick, S., and Akmal, M., 2005. Contradictory Trends in Global Income Inequality: A Tale of Two Biases. *Review of Income and Wealth, 51*(2), 201-229. doi: http://dx.doi.org/10.1111/j.1475-4991.2005.00152.x

EUROSTAT, 2007. European Union Statistics on Income and Living Conditions. from http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions

Ghosh, M., Nangia, N., and Kim, D. H., 1996. Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach. *Journal of the American Statistical Association, 91*(436), 1423-1431. doi: http://dx.doi.org/10.2307/2291568

Halley, R. M., 2004. Measures of Central Tendency, Location, and Dispersion in Salary Survey Research. *Compensation & Benefits Review, 36*(5), 39-52. doi: http://dx.doi.org/10.1177/0886368704268598

Kloek, T., and van Dijk, H. K., 1978. Efficient estimation of income distribution parameters. *Journal of Econometrics, 8*(1), 61-74. doi: http://dx.doi.org/10.1016/0304-4076(78)90090-8

Levy, P. S., and Lemeshow, S., 2008. *Sampling of Populationas. Methods and Applications* (4th ed. ed.). Hoboken: Wiley and Sons. doi:http://dx.doi.org/10.1002/9780470374597

Lohr, S. L., 2010. *Sampling: Design and Analysis* (2nd ed. ed.). Boston: Brooks/Cole.

Piegorsch, W. W., 2015. *Statistical Data Analysis. Foundations for Data Mining, Informatics, and Knowledge Discovery*. Chichester: Wiley and Sons.

Sala-i-Martin, X., 2006. The World Distribution of Income: Falling Poverty and … Convergence, Period*. *The Quarterly Journal of Economics, 121*(2), 351-397. doi: http://dx.doi.org/10.1162/qjec.2006.121.2.351

Terek, M., 2016. Odľahlé dáta a charakteristiky polohy v analýzach miezd a príjmov. *Revue sociálno-ekonomického rozvoja : vedecký recenzovaný on-line časopis, 2*(1), 114-126.

Tosenovsky, J., and Noskievicova, D., 2000. *Statisticke metody pro zlepsovani jakosti*. Ostrava: Montanex.

Wang, X., and Woo, W. T., 2011. The Size and Distribution of Hidden Household Income in China. *Asian Economic Papers, 10*(1), 1-26. doi: http://dx.doi.org/10.1162/ASEP_a_00064

Wonnacott, T. H., and Wonnacott, R., 1984. *Statistics for Business and Economics*. New York: Wiley and Sons.

Notes

[1] More in details, see in: European Union Statistics on Income and Living Conditions (EU-SILC), available at EUROSTAT (2007).
[2] Domain can be defined as subpopulation.
[3] We shall define an outlier in a set of data to be an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data (Barnett and Lewis, 1994, p. 7).
[4] More in details see in Halley (2004, pp. 39-52).
[5] More in details see in Terek (2016).
[6] More in details see for example in Terek (2016).
[7] Alternatively the term design weights is used.
[8] In fact, it is an empirical distribution of the observation from the population.
[9] More in details see for example in Tosenovsky and Noskievicova (2000).
[10] More in details see for example in Levy and Lemeshow (2008).
[11] For more details, see in Cochran (1977, pp. 8-9).

Copyright