

# OBTAINING THE INFORMATION ABOUT INCOMES FROM EU-SILC DATA AND MARKET ANALYSIS

Milan Terek

School of Management in Trencin, Slovakia

## ABSTRACT

The main goal of the paper is to present the methodology of collecting the detailed information about incomes from the EU-SILC survey useful for market analyses. The application of methods will be illustrated in the EU-SILC 2014 and 2016 data on the Slovak Republic. The population histogram, median, medial, and a measure of inequality based on median and medial of the whole gross household incomes for the whole Slovak Republic and separately for eight Slovak regions are estimated and compared.

**Keywords:** estimated median and medial incomes, estimated population histogram, estimated inequality of incomes

DOI: <http://dx.doi.org/10.15549/jeecar.v6i2.313>

## INTRODUCTION

An essential step in a firm's entry on the market is its analysis. The goal of the market analysis is to provide the basis for marketing decision making and planning. In the market analysis the most detailed information about the income structure of potential customers and its evolution is very important for the company producing and/or selling consumer goods and services. Besides the average income, it is also useful to know the regional income structure and the values of other characteristics, such as median, medial and income concentration. Such information is not routinely available on the sites of National Statistical Offices.

This paper shows how such information can be gained from the European Union Statistics on Income and Living Conditions Survey (EU-SILC) data. The survey is yearly executed in all EU

countries. Every year on its website the Statistical Office of the Slovak Republic publishes some results of the survey based on the processing of the collected data. The reports also include the information on average household income by a region, but not the information on incomes structure, median and medial incomes neither on income concentration. In the Slovak Republic, EU-SILC microdata are not freely available, but they can be obtained from the Statistical Office. We got the microdata for 2014 and 2016. These data will serve for the illustration of the methodology, allowing the collection of listed information that may be of great importance in marketing decisions and business planning. Moreover, the decisions on the regional structure of corporate activities can be based on the comparisons of the regions, and thus the paper presents such comparisons as well.

The EU-SILC data are the data from a complex survey. The complex survey is a statistical survey containing several components, such as random sampling, stratification, clustering, and so on. In such surveys the observations are not independent and (or) identically distributed, so the current methods of statistical inference and current software are not applicable in these cases or their application is limited. The use of sampling weights in estimation of population quantities is of interest. The estimation of population histogram, median, and medial are described in the paper. Construction of one measure of inequality based on median and medial is possible and is also presented and applied. The analysis of the regional structure of incomes in the Slovak Republic based on the above-mentioned estimated quantities is executed for the Slovak Republic as a whole and separately for its each region in the years 2014 and 2016. Then the results are compared.

### LITERATURE REVIEW

There many papers related to the distribution of income estimation and structure analysis, presenting these issues from different points of view. The overview of studies dealing with the estimation of income distribution parameters, the world distribution of income, the income inequality, the evolution of the highest incomes, etc. are listed in Terek (2017b).

In OECD Data (2018), the income inequality of OECD countries including the Slovak Republic in 2014 – 2017 is available. Some international comparisons are made in Growing (2008). However, no regional structure of incomes and regional income inequality is studied in the listed publications.

The question of how to analyze the regional structure of income in the Slovak Republic is analyzed in several studies. In Terek (2017a) the regional income structure analysis based on the estimated median in the year 2014 is executed. The structure of income in the year 2014 based on the estimated median, medial, and population histogram of the whole gross household income is analyzed in Terek (2017b), and in Terek and Muchova (2017), the estimated frequency distributions of the whole gross household income for the whole Slovak Republic and separately for all Slovak regions in 2014 is

analyzed. The measure of income inequality based on median and medial is described and applied in the analysis of the household incomes in Slovak regions in 2014 in Terek and Muchova (2018).

### METHODS

If the distribution of the analyzed data is skewed and/or outliers are present, the mean is not considered as the convenient measure for representation of a “typical” value. Distributions of income or wages are obviously skewed and outliers are present. An outlier in a set of data is an observation (or subset of observations) that appears to be inconsistent with the remainder of that set of data (Barnett and Lewis, 1994). Then, the interpretation power of the mean is very small (Halley, 2004). Generally, the median is considered to be a good measure for representing a “typical” value in such distributions because of its stability (it is insensitive toward the changes in frequencies of variable values) and robustness toward outliers. In the reports of the analyses of income or wages, the median is frequently presented along with the mean. There are known also other less-traditional measures of location enabling the characterization of the center of distribution and the representation of the typical income as well. For example, the use of trimmed mean, Winsorized mean or M-estimators is also possible. The calculation of median of the original data, the identification and removal of outliers by some of the known robust methods and calculating the mean from reduced data set are also possible. For more details see, for example, Terek and Tibensky (2014).

The data we analyze come from the complex survey in which the observations are not independent and identically distributed. The data sets presented by the Statistical Office also include sampling weights that can be used in correct estimation of the population quantities.

### Sampling Weights and Estimating a Probability Mass Function and Cumulative Distribution Function

The “model-free” or “distribution-free” approach to a sample survey is under consideration (for more details see Cochran, 1977); thus, only finite populations are

considered, and the probability mass function and cumulative distribution function can be defined exclusively for an observation from the finite population. They are defined as follows (Lohr, 2010): a value of probability mass function (PMF) in  $x$  is:

$$p(x) = \frac{N_{(x)}}{N}, \quad (1)$$

where  $N_{(x)}$  is the number of units whose value is  $x$  (symbol  $x$  denotes the variable under the study and also its values),

$N$  – size of finite population  $U = \{1, 2, \dots, N\}$

A value of cumulative distribution function (CDF) in  $x$  is:

$$F(x) = \sum_{y \leq x} p(y), \quad (2)$$

The probability mass function  $p(x)$  can be estimated by the empirical probability mass function  $\hat{p}(x)$ :

$$\hat{p}(x) = \frac{\sum_{i \in S: x_i = x} w_i}{\sum_{i \in S} w_i}, \quad (3)$$

where  $w_i$  is the sampling weight of unit  $i \in U$ ,

$S$  – sample from the population (subset containing  $n$  units from  $U$ ).

A sampling weight is the reversed value of the inclusion probability, where the inclusion probability is the probability that unit  $i \in U$  will occur in random sample. Sampling weight of unit  $i$  can be interpreted as the number of units in the population represented by unit  $i$ . These weights can be modified with regard to nonresponse and coverage error (for more details see Levy & Lemeshow, 2008).

The cumulative distribution function  $F(x)$  can be estimated by the empirical cumulative distribution function  $\hat{F}(x)$ :

$$\hat{F}(x) = \sum_{y \leq x} \hat{p}(y), \quad (4)$$

Based on sampling weights, the estimation of population quantities is possible.

If sampling weights are equal for all observation units, the sample is called self-weighting; if they are not, it is called non-self-weighting. If the sample is self-weighting, the point estimates can be calculated by standard procedures of statistical inference, supposing the statistical independence and identic distribution of observations, and by using a standard statistical software (the use of estimators based on sampling weights is not needed).

If the sample is non-self-weighting, the use of estimators based on sampling weights is necessary to capture the structure of data. In that case the point estimates produced by standard statistical software will be biased. The EU-SILC sample is non-self-weighting. The use of estimators based on sampling weights is appropriate.

### Frequency Distribution Estimation and Plotting Data from a Complex Survey

Below, the description of the way to utilize the sampling weights in population frequency distribution estimation is presented. If a sample is self-weighting, the frequencies calculated from sample data estimate the population frequency distribution. If a sample is non-self-weighting, the use of sampling weights in construction of a frequency distribution estimating the population frequency distribution is needed. The frequency of class  $j$  estimate is

$$\hat{n}(j) = \sum_{i \in S} w_i u_i(j), \quad (5)$$

where  $u_i(j) = 1$  if observation  $i$  is in class  $j$  and 0 otherwise (see also Terek & Muchova, 2017).

The relative frequency of class  $j$  estimate is

$$\hat{f}(j) = \frac{\hat{n}(j)}{\sum_{i \in S} w_i}, \quad (6)$$

In the case of a non-self-weighting sample, 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

$$\text{Height}(j) = \frac{\sum_{i \in S} w_i u_i(j)}{b \sum_{i \in S} w_i}, \quad (7)$$

The denominator in formula (7) ensures that the total area under the histogram equals 1. Such heights are sometimes called the densities of

relative frequencies (Wonnacott & Wonnacott, 1984).

### Estimating Some Population Quantities

Based on the empirical probability mass function  $\hat{p}(x)$  or empirical cumulative distribution function  $\hat{F}(x)$ , any population quantity can be estimated.

The population quantiles  $Q_p$  are of interest in the presented application. Let  $y_1$  be the largest value in the sample for which  $\hat{F}(y_1) \leq p$  and let  $y_2$  be the smallest value in the sample for which  $\hat{F}(y_2) \geq p$ . Then an estimated population quantile  $\hat{Q}_p$  is:

$$\hat{Q}_p = y_1 + \frac{p - \hat{F}(y_1)}{\hat{F}(y_2) - \hat{F}(y_1)}(y_2 - y_1), \quad (8)$$

Medial ( $Ml$ ) is such a value for which the sum of variable values less or equal to  $Ml$  is equal to the half of the variable total. If all the values of variable are nonnegative, then medial is greater or equal to median  $Q_{0.5}$  (Dagnelie, 1998). In estimation of medial the empirical probability mass function  $\hat{p}_\tau(x)$  and empirical cumulative distribution function  $\hat{F}_\tau(x)$  defined as:

$$\hat{p}_\tau(x) = \frac{\sum_{i \in S: x_i = x} w_i x_i}{\sum_{i \in S} w_i x_i}, \quad (9)$$

$$\hat{F}_\tau(x) = \sum_{y \leq x} \hat{p}_\tau(y), \quad (10)$$

can be used (the detailed explanation of their formulation is in Attachment 1).

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

$$\hat{Ml} = y_1 + \frac{0.5 - \hat{F}_\tau(y_1)}{\hat{F}_\tau(y_2) - \hat{F}_\tau(y_1)}(y_2 - y_1), \quad (11)$$

The proportion of units having values less than or equal to medial can be estimated with the help of  $\hat{F}_\tau(\hat{Ml})$  as a special case of the mean for a variable taking two values:  $x_i = 1$ , if unit  $i$  has the value less or equal to  $\hat{Ml}$  and  $x_i = 0$ , if unit  $i$  has

the value greater than the estimated medial. The detailed explanation is in Attachment 2.

### Measures of Inequality

Inequality (concentration) in the broadest sense means the accumulation of the objects, for example, income units, to subjects, for example, households (Marfels, 1971). The measures of inequality are used mainly for the measurement of the distribution of the wages or income totals. If, for example, some percentage of households obtains the same total of incomes, the distribution of incomes is perfectly uniform the inequality is null. If one household obtains the whole income total, the inequality is maximal. In general, three measures of inequality are known – Lorentz curve, Gini index, and the measure of inequality based on medial (Coeurjolly, 2015). The last one is used in the application under consideration:

$$\Delta = \frac{Ml - Q_{0.5}}{R}, \quad (12)$$

where  $R$  is the range. It is estimated by:

$$\hat{\Delta} = \frac{\hat{Ml} - \hat{Q}_{0.5}}{\hat{R}}, \quad (13)$$

where  $\hat{R}$  is the range in the sample.

The values of  $\hat{\Delta}$  can be negatively influenced by outliers. The quantities  $\hat{Ml}$  and  $\hat{Q}_{0.5}$  are robust toward the outliers, thus the outliers are determined and removed from the data set only in  $\hat{R}$  calculation process. The Tukey method of outliers detection based on interquartile range is used. The values more than 1.5 of interquartile range below the first quartile and 1.5 of interquartile range above the third quartile are considered to be outliers. The quartiles needed for interquartile range calculation are estimated according to (8), for  $p = 0.25$  and  $p = 0.75$ .

### REGIONAL STRUCTURE OF INCOMES AND ITS EVOLUTION

The analysis of regional structure of incomes was realized on the data from the EU-SILC survey executed in the Slovak Republic in 2014 and 2016. The stratified two-stage survey is applied in the Slovak Republic on a regular yearly basis. Two stratification variables are used in the survey, i.e. region and settlement size. Eight regions in the Slovak Republic include Bratislava,

Trnava, Trencin, Nitra, Zilina, Banska Bystrica, Kosice, and Presov. 5,490 households in 2014 and 5,738 in 2016 with sampling weights were introduced to the databases and analyzed in the paper. For the whole Slovak Republic and separately for each region the values of the empirical probability mass function were calculated by (3), the values of the empirical cumulative distribution function by (4).

### Estimation of Frequency Distributions and Population Histograms

In Table 1 are estimated frequencies for the whole Slovak Republic calculated by (5). The tables with the estimated frequencies for each

region can be constructed in the same way. Alternatively, the histograms estimating the population histograms can be constructed.

The histograms estimating the corresponding population histograms of the households' distribution by the whole gross income in the years 2014 and 2016 were constructed for the Slovak Republic as a whole and for each region separately. The heights (densities of relative frequencies) in histograms were calculated by (7). The widths of classes in histograms are  $b = 10,000$  euros. The area of a rectangle in a histogram is equal to relative frequency.

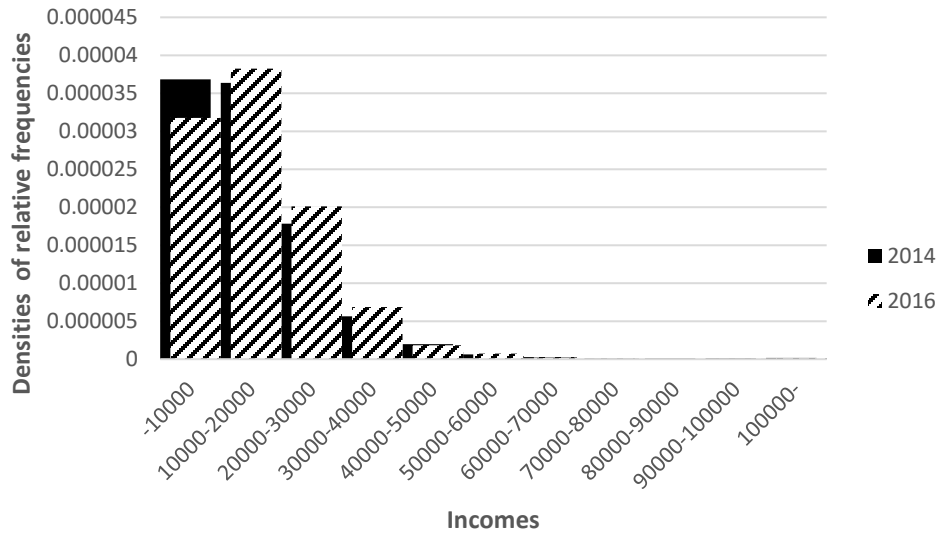
**Table 1.** Distribution estimating distribution of Slovak households by the whole gross household income in 2014 and 2016

Income (Euros)	Estimated frequency in 2014	Estimated frequency in 2016
-10 000	681,765	589,299
10 000-20 000	673,063	707,783
20 000-30 000	329,995	372,783
30 000-40 000	104,565	126,550
40 000-50 000	36,541	33,824
50 000-60 000	11,861	13,532
60 000-70 000	5,038	4,341
70 000-80 000	1,381	1,088
80 000-90 000	1,647	306
90 000-100 000	1,791	681
100 000-	3,195	1,873
Total	1,850,842	1,852,060

Source: Developed by author.

The histogram estimating population histogram for the Slovak Republic in 2014 and 2016 is presented in Figure 1. The black rectangles represent the year 2014, hatched ones, the year 2016 on the following figures. Because of the possibility of direct comparison of histograms, the rectangles in histograms do not

touch. Figure 1 shows that the proportion of households with incomes less or equal to 10,000 euros decreased in 2016 compared to 2014. On the other hand, proportions of households having incomes from 10,000 to 40,000 euros increased.



**Figure 1.** Histograms of the Slovak households' distribution by the whole gross income in 2014 and 2016.

Source: Developed by author.

Figure 2 shows that in region Bratislava the proportion of “poorest” households decreased in 2016 compared to 2014 and the proportion of households having incomes greater than 10,000 euros and less or equal to 30,000 euros significantly grew. The histogram also shows another change, namely the proportion of the “richest” households being lower in 2016 than the one in 2014.

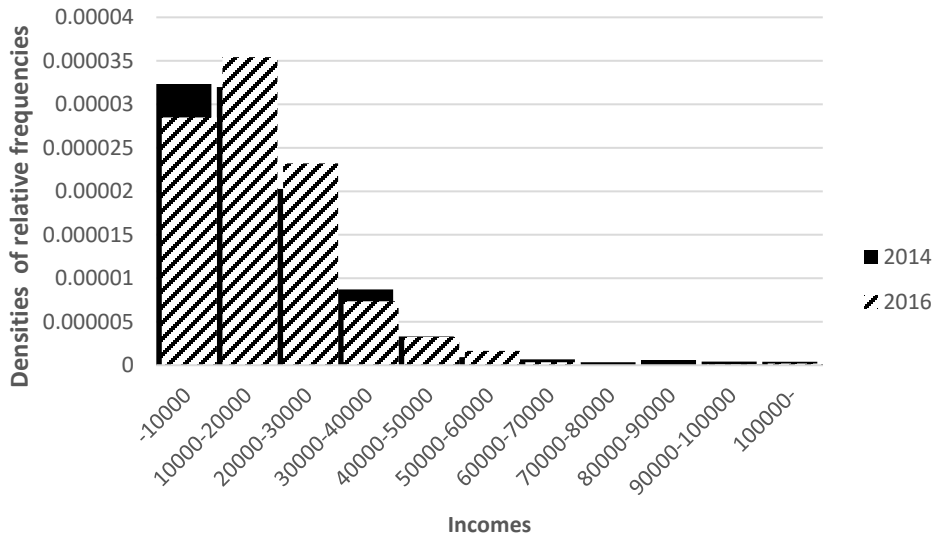
In region Trnava (Figure 3) there is a significant increase of proportion incomes higher than 30,000 euros and less or equal to 60,000 euros which seems interesting.

Region Trencin (Figure 4) differs from the previous two regions, by the decrease the proportion of the households with incomes

higher than 10,000 euros and less or equal to 20,000 euros, in 2016. The higher proportions in the interval from 20,000 to 50,000 euros in 2016 seem as interesting.

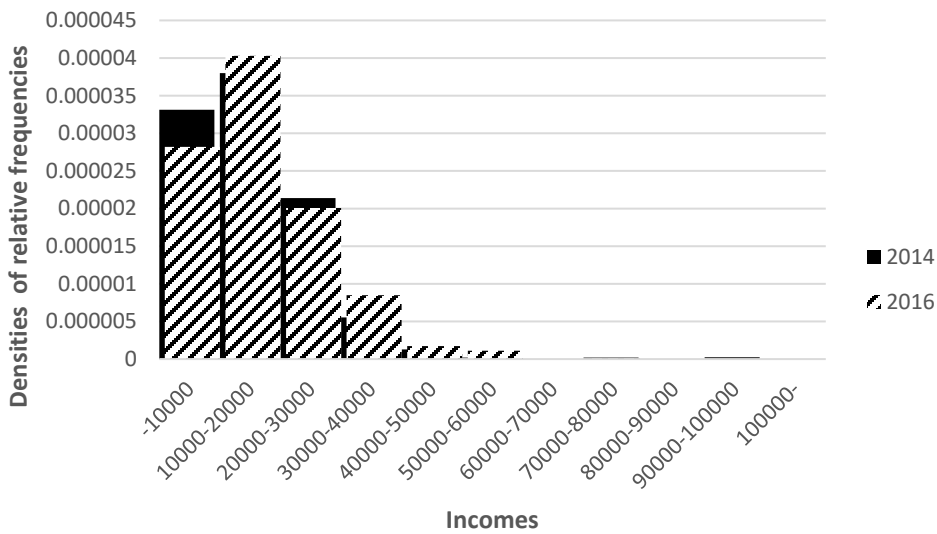
The development of incomes structure in region Nitra (Figure 5) is very close to the Slovak Republic incomes structure development. The number of the “poorest” households decreased in comparison with the “richer” ones.

The incomes development in Zilina region (Figure 6) does not differ very much from the Slovak Republic development. The proportions of the “richest” households (with incomes higher than 40,000 euros) is decreasing which we find very interesting.



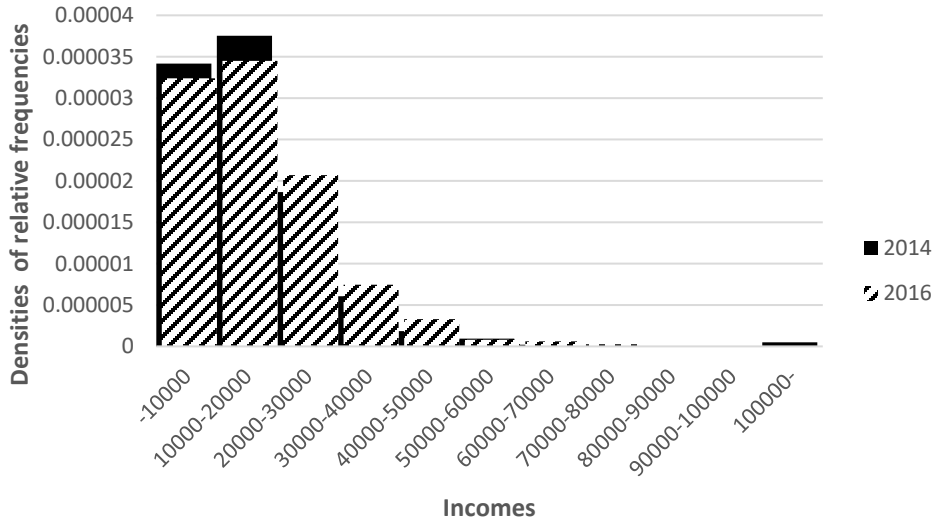
**Figure 2.** Histograms of Bratislava region households' distribution by the whole gross income in 2014 and 2016.

Source: Developed by author.



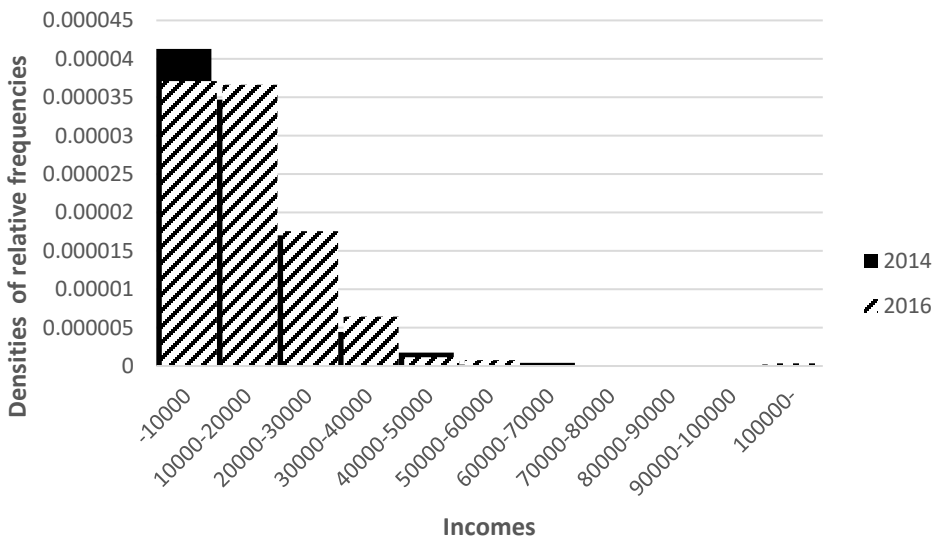
**Figure 3.** Histograms of Trnava region households' distribution by the whole gross income in 2014 and 2016

Source: Developed by author.



**Figure 4.** Histograms of Trencin region households' distribution by the whole gross income in 2014 and 2016.

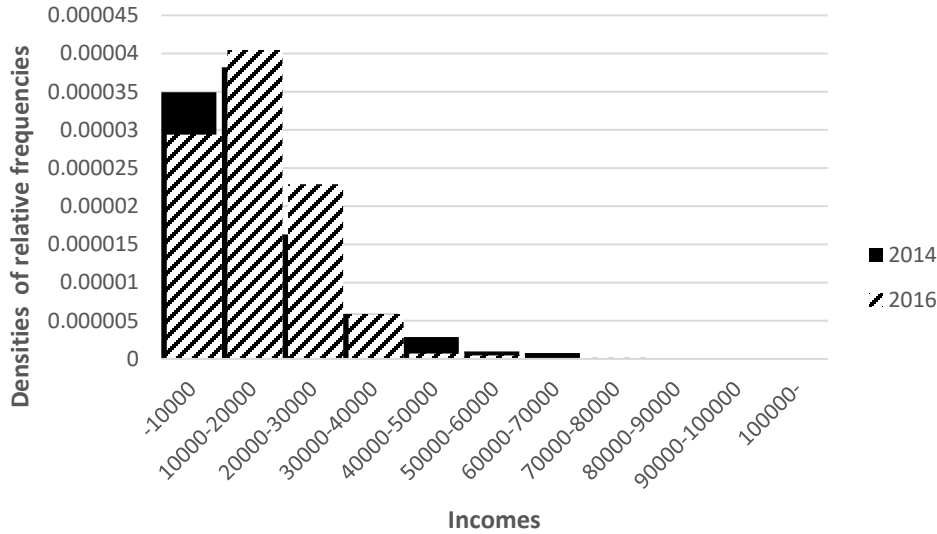
Source: Developed by author



**Figure 5.** Histogram of Nitra region households' distribution by the whole gross income in 2014 and 2016.

Source: Developed by author



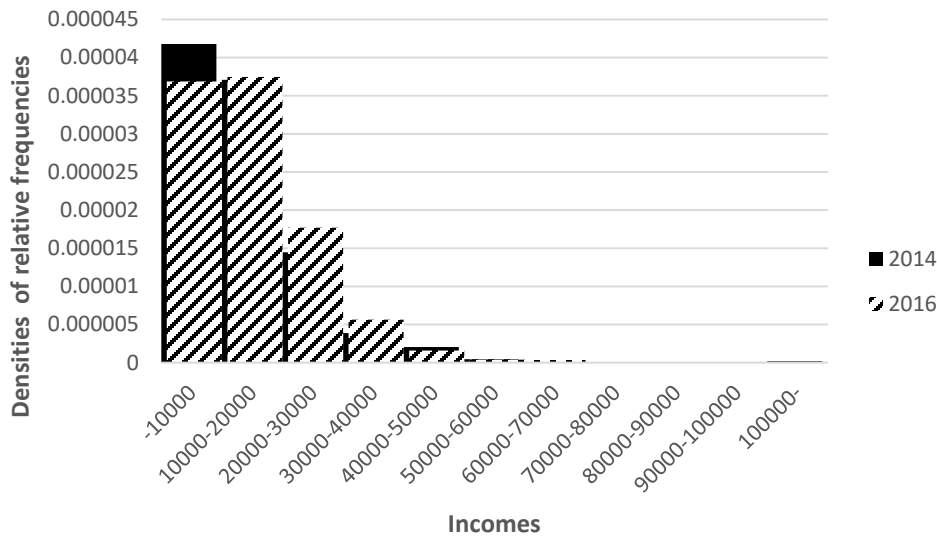


**Figure 6.** Histograms of Zilina region households' distribution by the whole gross income in 2014 and 2016.

Source: Developed by author/

The incomes development in Banska Bystrica region (Figure 7) also maps the Slovak Republic development. The increase in the proportions of households with their incomes in the interval between 10,000 – 20,000 euros differs from

other regions only moderately. The proportions of households having incomes in the interval between 20,000 – 40,000 euros increased more significantly.

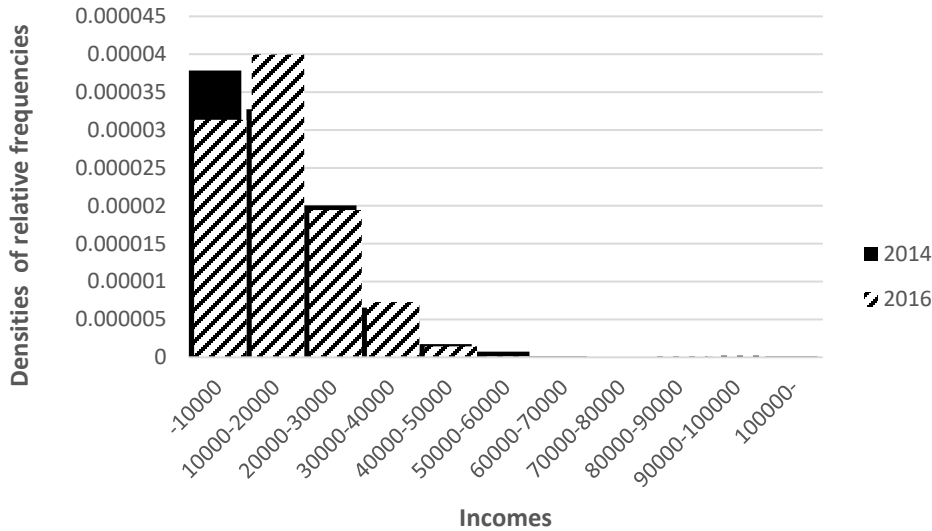


**Figure 7.** Histograms of Banska Bystrica region households' distribution by the whole gross income in 2014 and 2016.

Source: Developed by author

The great increase in the proportion of households with incomes 10,000 to 20,000 euros in Presov region (Figure 8) we find interesting.

The changes in the proportions of the “richer” households are only moderate.

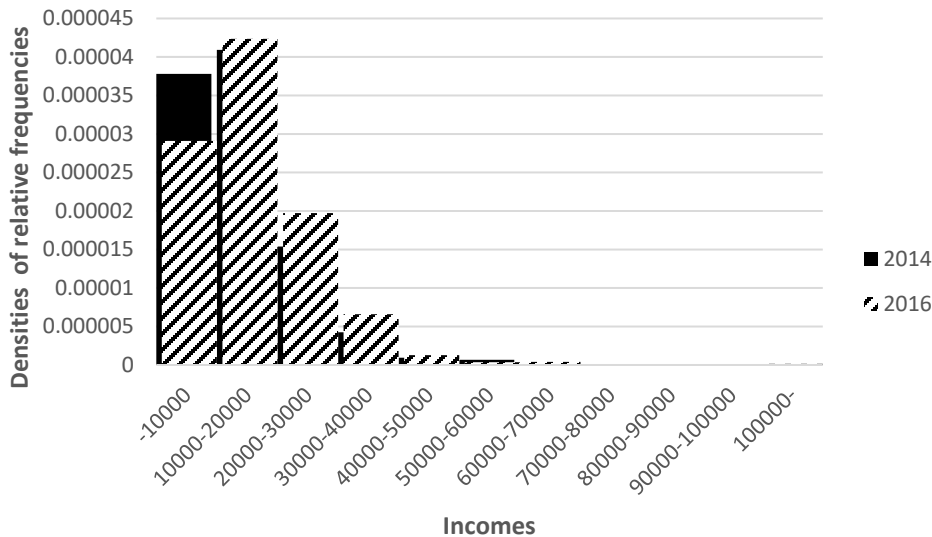


**Figure 8.** Histogram of Presov region households' distribution by the whole gross income in 2014 and 2016.

Source: Developed by author.

Figure 9 shows the increase in the proportions of households with incomes in the interval from

10,000 to 50,000 euros in Kosice region.



**Figure 9.** Histogram of Kosice region households' distribution by the whole gross income in 2014 and 2016.

Source: Developed by author.

### Estimates of Median Incomes

The estimate of median whole gross household income was calculated according to (8) for the Slovak Republic as a whole and separately for each region. The estimate of population median whole gross household income for the Slovak Republic in the year 2014 equals 13,305.83 euros

and in the year 2016 equals 14,631.33 euros (see also Terek, 2018). The estimates of median in regions are presented in Table 2. In the table in brackets are the serial numbers of regions according to the median whole gross household income in the corresponding year.

**Table 2.** Regional structure of median whole gross household income in 2014 and 2016

Region number	Region name	Estimate of median whole gross household income in 2014 (Euros)	Estimate of median whole gross household income in 2016 (Euros)
1	Bratislava	14,491 (1.)	16,000 (1.)
2	Trnava	13,969 (4.)	15,258 (3.)
3	Trencin	14,368 (2.)	15,557 (2.)
4	Nitra	12,380 (7.)	13,286 (7.)
5	Zilina	14,055 (3.)	15,023 (4.)
6	Banska Bystrica	11,746 (8.)	12,900 (8.)
7	Presov	13,595 (5.)	14,372 (6.)
8	Kosice	13,118 (6.)	14,678 (5.)

Source: Developed by author.

### Estimates of Medial Incomes and Proportions of Households with Incomes Less or Equal to Medial

The values of empirical probability mass function  $\hat{p}_r(x)$  were calculated by (9) and the values of empirical cumulative distribution function  $\hat{F}_r(x)$  by (10) for the Slovak Republic as a whole and separately for each region. The estimate of medial whole gross household income was calculated by (11) for the whole Slovak Republic and separately for each region. The estimated proportion of households with incomes less than or equal to medial was calculated as  $\hat{\mu}_K$  by (10) (in Attachment 2).

The estimate of medial whole gross household income in 2014 for the whole Slovak Republic accounts for 20,356 euros and the estimated

proportion (in percent) of households with incomes less or equal to medial is 74.15%. A half of income total in the Slovak Republic in 2014 was distributed among 74.15 % of “poorer” households (with income less or equal to 20,356 euros) and the other half among the remaining 25.85 % of “richer” households (with income higher or equal to 20,356 euros). In 2016 a half of income total was distributed among 72.62 % of “poorer” households (with income less or equal to 21,025 euros) and the other half among the remaining 27.38 % of “richer” households (with income higher or equal to 21,025 euros). The results for individual regions are presented in Table 3. The numbers in the table (in brackets) show their order by the medial whole gross household income in the corresponding year.

**Table 3.** Regional structure of medial whole gross household income in 2014 and 2016.

Region number	Region name	Estimate of medial whole gross household income in 2014 (Euros)	Estimated proportion of households with incomes less or equal to medial in 2014 (%)	Estimate of medial whole gross household income in 2016 (Euros)	Estimated proportion of households with incomes less or equal to medial in 2016 (%)
1	Bratislava	24,875 (1.)	76.71	24,291 (1.)	73.15
2	Trnava	20,331 (5.)	72.44	21,000 (3.)	71.95
3	Trencin	20,556 (2.)	73.40	22,003 (2.)	72.88
4	Nitra	19,364 (6.)	74.61	20,089 (6.)	74.02
5	Zilina	20,492 (3.)	73.88	20,349 (4.)	70.27
6	Banska Bystrica	18,302 (7.)	74.88	19,619 (8.)	73.49
7	Presov	20,490 (4.)	73.34	20,290 (5.)	71.56
8	Kosice	18,194 (8.)	71.90	19,790 (7.)	70.98

Source: Developed by author.

#### Estimating the Inequality of Incomes

The values of measure of inequality  $\Delta$  defined by (12) were estimated by  $\hat{\Delta}$  according to (13).

The value of  $\hat{\Delta}$  in the whole Slovak Republic was 17.46 % in 2014 and 15.32 % in 2016. The results for the regions are shown in Table 4.

**Table 4.** Regional structure of inequality of the whole gross household income in 2014 and 2016.

Region number	Region name	2014 $\hat{\Delta}$ (%)	2016 $\hat{\Delta}$ (%)	Decrease of $\hat{\Delta}$ (percentage points)
1	Bratislava	21.90	16.95	4.95
2	Trnava	15.86	13.67	2.19
3	Trencin	15.51	14.89	0.62
4	Nitra	17.98	17.09	0.89
5	Zilina	15.97	13.36	2.61
6	Banska Bystrica	18.45	16.97	1.48
7	Presov	16.98	15.14	1.84
8	Kosice	14.82	13.87	0.95

Source: Developed by author

### CONCLUSIONS

The results can serve well in estimating purchasing power and potential demand across the country and its regions. In addition, they may be useful in customer profiling and in some sociological studies.

In general, the median provides a better idea of "typical value" as it is understood by most public. The statement "a half of the households in the Slovak Republic had a whole gross household income less or equal to 14,631.33 euros in 2016" provides a better idea of typical income than a statement "average whole gross household income was 16,992 euros in 2016 (EU-SILC, 2017)". In addition, a relatively large difference between the median and the average income is due to the difference in robustness of these two characteristics to extreme data. The arithmetic mean is simply not a good characteristic of a "typical value" in significantly skewed distributions. All listed histograms account for significantly skewed distributions. Of course, this is not a specific feature of the Slovak Republic position, it is a general feature of incomes and wages. Changes in the positioning of regions can be used in marketing planning, even though, in the Slovak Republic the changes in 2016 compared to 2014 are not large.

The medial provides information on income that divides the income total into two halves. A half of the income total is distributed among households with the income less or equal to medial. Being aware of the medial and especially monitoring its changes on the regional level can be particularly useful in marketing planning.

In the Slovak Republic, the regional differences in the medial whole gross household incomes are not large and they decreased in 2016 compared to 2014. The differences in ranking regions in 2016 compared to 2014 are only moderate. The findings showing that a half of the incomes total is distributed among 71.90 – 76.71 % of "poorer" households and the other half among the rest of "richer" households in all Slovak regions in 2014, and in 2016 the situation is not very different (a half of the incomes total is distributed among 70.27 – 74.02 % of "poorer" households and the other half, among the rest of "richer" households). This can also be interesting for marketing decision making.

With median and medial, the calculation of the corresponding measure of inequality is simple. In the Slovak Republic the inequality of incomes decreased in the whole Republic (from 17.46 % to 15.32 %) as well as in its regions in 2016 compared to 2014. This information is also useful in marketing decision making.

It must be noted that sampling design of EU-SILC may vary from country to country. Different stratification variables can be used as well as a variety of degrees within a multi-stage sampling.

### ACKNOWLEDGEMENTS

The paper was supported by grant from Grant Agency of VEGA no. 1/0393/16 entitled "European Union in Post Crisis Period – Macro and Microeconomic Aspects."

### REFERENCES

- Barnett, V., and Lewis, T. (1994). *Outliers in Statistical Data*, Hoboken: Wiley and Sons
- Cochran, W. G. (1977). *Sampling Techniques*, New York: J. Wiley and Sons
- Coeurjolly, J.-F. (2015). Chapitre 2. Caractéristiques des distributions à une variable quantitative, from [http://www-ljk.imag.fr/membres/Jean-Francois.Coeurjolly/documents/L1/chap2\\_int.pdf](http://www-ljk.imag.fr/membres/Jean-Francois.Coeurjolly/documents/L1/chap2_int.pdf)
- Dagnelie, P. (1998). *Statistique Théorique et Appliquée. Tom 1 – Statistique Descriptive et Bases de l'Inférence Statistique*, Paris: DeBoeck and Larcier
- EU SILC 2017, from [https://slovak.statistics.sk/wps/portal/a778c769-b8f6-419a-86e0-0ccf04240498!/ut/p/z0/jY27CglxFES\\_xclYe29izKNcFUTBQkRY00gSdjXuUze-\\_t61s7QZZg4cBgxkYBr7CCcbQ9vYatgHI45buVKzGU0R5XSBq\\_V-sV3uNEWOsAYTXJ08fZ1gQpmSgjPKlWaToX91dtvMNycwnY1nEpqihcxKqbwUmjhVCMKptkSJHAI6XyBnHLIWXzVcrleTgvFtE\\_NXhKx1vT2TviTd3Y1xiCqU1od8jl8-j-X7l\\_x10pXGvZ\\_p6APQ1jV6/](https://slovak.statistics.sk/wps/portal/a778c769-b8f6-419a-86e0-0ccf04240498!/ut/p/z0/jY27CglxFES_xclYe29izKNcFUTBQkRY00gSdjXuUze-_t61s7QZZg4cBgxkYBr7CCcbQ9vYatgHI45buVKzGU0R5XSBq_V-sV3uNEWOsAYTXJ08fZ1gQpmSgjPKlWaToX91dtvMNycwnY1nEpqihcxKqbwUmjhVCMKptkSJHAI6XyBnHLIWXzVcrleTgvFtE_NXhKx1vT2TviTd3Y1xiCqU1od8jl8-j-X7l_x10pXGvZ_p6APQ1jV6/)
- European Union Statistics on Income and Living Conditions (EU-SILC), from

- <http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions>
- Growing Unequal? Income Distribution and Poverty in OECD Countries, 2008, from <https://www.mzv.sk/documents/10182/12385/Growing+Unequal.pdf>
- Halley, R. M. (2004). "Measures of Central Tendency, Location, and Dispersion in Wage Survey Research", *Compensation and Benefits*, 2004/36, 39 (2004) pp. 39-52
- Levy, P. S., and Lemeshow, S. (2008). *Sampling of Populations. Methods and Applications*, Fourth Edition, Hoboken: Wiley and Sons
- Lohr, S. L. (2010). *Sampling: Design and Analysis*, 2nd edition, Boston: Brooks/Cole
- Marfels, Ch. (1971). "Absolute and Relative Measures of Concentration Reconsidered", *Kyklos. International Review for Social Sciences*, Volume 24, Issue 4, pp. 753-766
- OECD Data, *Income Inequality*, 2018, from <https://data.oecd.org/inequality/income-inequality.htm>.
- Terek, M., and Tibensky, M. (2014). "Outliers and Some Non-Traditional Measures of Location in Analysis of Wages", *European Scientific Journal*, September 2014, Special Edition, Vol. 1, ISSN 1857 - 7881
- Terek, M. (2017a). *Regional Incomes Structure Analysis Based on Complex Survey Data*. 16th Conference on Applied Mathematics APLIMAT 2017 Proceedings. Bratislava: Spektrum STU, 2017. ISBN 978-80-227-4650-2, s. 1504 - 1512
- Terek, M. (2017b). *Regional Incomes Structure Analysis in Slovak Republic on the Basis of EU-SILC Data*. *Scientific Annals of Economics and Business*, 2017, VOL 64, NO. 2
- Terek, M., and Muchova, E. (2017). *The Structure of Incomes Analysis in Slovak Republic and Regions of the Slovak Republic Based on EU-SILC Data*. *International Journal of Economic Research*, Volume 14, Number 20, 2017, pp. 425 - 434
- Terek, M., and Muchova, E. (2018). *Regional Differences in Income Distribution in the Slovak Republic*. *Proceedings from 5th 2018 IBSM - International Conference on Business, Management and Accounting*, April 19-21, 2018, Hanoi, Vietnam. CAAL International Education, Training and Consulting. ISBN 978-602-72911-6-4
- Terek, M. (2018). *Analysis of Incomes Structure in Slovak Republic in the year 2016*. *Conference Proceedings, Second International Scientific Conference on IT, Tourism, Economics, Management and Agriculture - ITEMA 2018*, November 8, 2018 Graz University of Technology, Graz, Austria. Belgrade: Association of Economists and Managers of the Balkans, 2018, ISBN 978-86-80194-13-4, from <http://www.itema-conference.com/draft-conference-proceedings.html>
- Wonnacott, T. H., and Wonnacott, R. J. (1984). *Statistics for Business and Economics*. New York: Wiley and Sons (French translation (1990), *Statistique*, Paris: Economica).

### Attachment 1

#### Empirical probability mass function and cumulative distribution function of class total

The relations enabling the estimation of medial using sampling weights will be formulated. The sum of variable values for all observations taking on the value  $x$  is called the class total of  $x$ . The medial is calculated as the median, but on the basis of class totals instead of frequencies. A value of probability mass function (PMF) in  $x$  is

$$p_{\tau}(x) = \frac{\tau(x)}{\tau}$$

where  $\tau(x)$  is the class total of  $x$  and  $\tau$  is the total of variable values in the population. A value of cumulative distribution function (CDF) in  $x$  is

$$F_{\tau}(x) = \sum_{y \leq x} p_{\tau}(y)$$

The probability mass function  $p_{\tau}(x)$  is estimated by empirical probability mass function  $\hat{p}_{\tau}(x)$  defined by the sum of products of the weights and variable values taking on the value  $x$  in the sample divided by the sum of the products of the weights and variable values in the sample:

$$\hat{p}_{\tau}(x) = \frac{\sum_{i \in S: x_i = x} w_i x_i}{\sum_{i \in S} w_i x_i}$$

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

$$\hat{F}_{\tau}(x) = \sum_{y \leq x} \hat{p}_{\tau}(y)$$

In our application,  $x$  is the value of income,  $i$  is the index of the household and  $x_i$  is the income of the  $i$ -th household. The probability  $p_{\tau}(x)$  can be interpreted as the probability that one euro randomly selected from all incomes is from the incomes of size  $x$  and  $F_{\tau}(x)$  as the probability that one euro randomly selected from all incomes is from the incomes less or equal to  $x$ .

### Attachment 2

#### Estimating of proportion by sampling weights

The proportion of units having a characteristic is simply a special case of the mean, obtained by letting  $x_i = 1$  if unit  $i$  has the characteristic of interest, and  $x_i = 0$  if unit  $i$  does not have the characteristic. In general, the mean  $\mu_K$  can be estimated with aid of sampling weights by (Lohr, 2010)

$$\hat{\mu}_K = \frac{\sum_{i \in S} w_i x_i}{\sum_{i \in S} w_i}$$

When  $x_i$  is 0 – 1 variable mentioned above, the proportion  $\pi_K$  in the population can be estimated by the same formula.

### ABOUT THE AUTHORS

Milan Terek, email: milan.terek1@gmail.com

**Dr. Milan Terek** is a full professor at the School of Management in Trencin (course leader on: Introduction to Statistics, Statistics, Quantitative Methods for Managers, Quantitative methods in Business Management Research). Before he worked at the University of Economics in Bratislava (course leader on: Statistics, Statistical Quality Control, Decision Analysis, Data Mining, Survey Sampling, Linear Programming, Nonlinear Programming, Operations Research, System Modeling). His research activities are oriented on the applications of statistical methods in economics and management.