ANALYSIS OF INCOMES STRUCTURE IN SLOVAK REPUBLIC IN THE YEAR 2016

Milan Terek[145]

Abstract: *The paper deals with the analysis of incomes structure in regions of Slovak republic based on European Union statistics on income and living conditions in Slovak republic data in the year 2016. The empirical probability mass function and empirical cumulative distribution function are formulated using the given sampling weights. Based on these functions the median whole gross household incomes for the whole Slovak republic and separately for eight Slovak regions are estimated. The results of analysis show that the differences among Slovak regions according to the median whole gross household incomes are mostly only moderate.*

Keywords: *median incomes, sampling weights, empirical probability mass function, empirical cumulative distribution function*

1. INTRODUCTION

The analysis of incomes structure in Slovak republic will be analyzed based on the data from the European Union Statistics on Income and Living Conditions (EU-SILC) realized in Slovak republic in the year 2016. 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 (More in details see in [1]). The start of the EU-SILC instrument was in 2004 for the EU-15.

The survey containing more components such as random sampling, stratification, clustering, ratio estimation and so on is obviously called complex survey. All these procedures require using of auxiliary information. When the survey is complex, sampling weights based on auxiliary information and design effects are commonly used to simplify matters and to provide the correct results.

The construction of empirical probability mass function and empirical cumulative distribution function enabling to estimate the median with aid of sampling weights will be described in the paper. The analysis of the regional structure of incomes in Slovak republic based on estimated median whole gross household incomes will be realized.

---

[145] Vysoka skola manazmentu/City University of Seattle Programs, Panonska cesta 17, 851 04 Bratislava, Slovak Republic

2. METHODS

Distributions of incomes or wages are obviously skewed and outliers are present. An outlier in a set of data is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data ([2], p. 7). In such distributions the median is considered as good measure of central tendency because of its stability and robustness toward outliers. Alternatively also some less-traditional measures of location could be interesting as good measures of central tendency in such distributions, for example trimmed mean, Winsorized mean or M-estimators (More in details see for example in [3]). The using of the mean after removing of outliers is also possible and offers good results (See for example in [4]).

Suppose we know the size N of finite population U. Symbol x denotes variable under study and also its values, $U = \{1, 2, ... N\}$ 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 inclusion probability − 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 arbitrary sampling designs can be expressed by sampling weights. If sampling weights are equal for all observations units, the sample is called self-weighting. Each observed unit in such sample represents the same number of unobserved units in the population. If sampling weights are not equal for all observations units, the sample is called non-self-weighting. If the sample is non-self-weighting, point estimates of quantities produced by standard statistical software will be biased.

Suppose that 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

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

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 (For more details see in [5], pp. 8-9).

Empirical probability mass function $\hat{p}(x)$ is defined as follows

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

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

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

Population quantiles $\hat{Q}_p$ are estimated as follows (According to [6]). Let y_1 be the largest value in the sample for which $\hat{F}(y_1) \leq p$ and let y_2 is smallest value in the sample for which $\hat{F}(y_2) \geq p$. Then

$$\hat{Q}_p = y_1 + \frac{p - \hat{F}(y_1)}{\hat{F}(y_2) - \hat{F}(y_1)}(y_2 - y_1) \tag{6}$$

3. ESTIMATES OF MEDIAN WHOLE GROSS HOUSEHOLD INCOMES IN REGIONS OF SLOVAK REPUBLIC

The analysis of regional structure of incomes was effectuated on the data from the survey EU-SILC realized in Slovak republic in 2016. The stratified two-stage survey is regularly yearly applied in Slovak republic. A stratification is effected with two stratification variables – region and settlement size. There are eight regions in Slovak republic: Bratislava, Trnava, Trencin and Nitra in western Slovakia, Zilina and Banska Bystrica in central Slovakia, Kosice and Presov in eastern Slovakia. The EU-SILC sample is non-self-weighting. The using of sampling weights in estimation is suitable.

The 5,738 households were included to databases in 2016 also with the corresponding sampling weights. Data from EU-SILC 2016 are concentrated in many sets. Each household has one identification number. The analysis of the whole gross household incomes in eight Slovak regions was realized. 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. Each region was analyzed separately. The values of the empirical probability mass function were calculated according to (4) and the values of the empirical cumulative distribution function according to (5) for each region.

The estimate of median whole gross household income was calculated according to relation (6) separately for each region and for the whole Slovak republic. The estimate of population median whole gross household income for the whole Slovak republic in the year 2016 equals 14,631 euro's. The estimates of median in regions are in Table 1. In the brackets are the order numbers of layout according to the median whole gross household income.

Table 1: Regional structure of median whole gross household income in 2016

Region number	Region name	Estimate of median whole gross household income in 2016 (Euros)
1	Bratislava	16,000 (1.)
2	Trnava	15,258 (3.)
3	Trencin	15,557 (2.)
4	Nitra	13,286 (7.)

5	Zilina	15,023 (4.)
6	Banska Bystrica	12,900 (8.)
7	Presov	14,372 (6.)
8	Kosice	14,678 (5.)

4. CONCLUSIONS

The obtained ordering of regions according to median whole gross household income is very interesting. The difference between first Bratislava and second Trencin regions is not very large. The median household incomes of the third Trnava, fourth Zilina and fifth Kosice 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 median whole gross household income of the first Bratislava is about 24 % greater than in the last Banska Bystrica. The first five regions have the median whole gross household income greater and the resting regions less than the whole Slovak republic.

If the sample is non-self-weighting, point estimates of means, totals and other quantities produced by standard statistical software will be biased. So in this case the using the standard statistical software to point estimation is inappropriate. It was the case also in mentioned application. The point estimation of parameters using sampling weights was needed. It is clear that the estimate obtained with aid of finite weights which allow the used sample design, non-response and potentially also coverage error better reflects the reality. Note that the steps of the weighting process are quite subjective. Even computing the base weights can be subjective; for example, in many telephone surveys, the number of telephones per household is not obtained and, thus, not used for computing inclusion probabilities. But there is general agreement in the statistical community that, for estimating descriptive statistics, weights should always be used even though some estimates may not benefit from this practice.

REFERENCES

[1] European Union Statistics on Income and Living Conditions (EU-SILC). Available at: http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions

[2] Barnett, V. – Lewis, T. (1994) *Outliers in Statistical Data,* Wiley and Sons, Hoboken, p. 7.

[3] Terek, M., Tibensky, M. (2014) Outliers and Some Non-Traditional Measures of Location in Analysis of Wages. *European Scientific Journal*, Special Edition, 2014, Vol. 1.

[4] Terek, M. (2016): Odlahle data a charakteristiky polohy v analyzach miezd a prijmov. In: *Revue socialno-ekonomickeho rozvoja*, Vol. II, number 1/2016.

[5] Cochran, W. G. (1977) *Sampling Techniques,* Wiley and Sons, New York.

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