BONFERRONI INDEX DECOMPOSITION AND THE SHAPLEY METHOD

Giovanni Maria Giorgi, Alessio Guandalini

Introduction

In the last years the Bonferroni inequality index (B, Bonferroni, 1930) has been analysed with more attention for its particular characteristics. The Bonferroni, as the Gini (1914) index, has been identified as special cases of a general formula by De Vergottini (1956). Further studies on B have been conducted by Piesch (1975) and Nygård and Sandström (1981). Recently, new and interesting interpretations and extensions of B have been proposed.

A widespread topic in the literature on inequality measures is their decomposition. Many contributions are related to the Gini index R (Gini, 1914). Tarsitano (1990) showed various standard results used to investigate the decomposition of B and Bárcena-Martin and Silber (2013) derived an algorithm that greatly simplifies it.

In the field of inequality index decomposition two main lines of research can be distinguished: decomposition by income sources and by population subgroups. The former is widely treated, whilst less attention has been paid on the latter (Giorgi, 2011). The reason lies in the difficult to decompose additively some inequality indices, such as R and B, by population subgroups. To overcome this drawback Deutsch and Silber (2007) used the so-called Shapley method on R.

In the present paper the same method has been applied on B. Several interesting similarities and differences among the two indices are highlighted. Furthermore, some properties of B have been object of deeper investigation.

The paper is organized as follows: in section 1 the original expression and the main properties of B are presented. In section 2, the Shapley method is quickly surveyed and a numerical illustration is provided. In section 4 the method is applied to real data (It-SILC data referred to 2009). Finally, conclusions and future prospects of research are discussed in section 5.

1. The Bonferroni inequality index

The expression of B proposed by Bonferroni (1930, p. 55 and p. 85) is a function of partial means:

$$B = \frac{1}{N-1}\sum_{i=1}^{N-1}\frac{(\mu-\mu_i)}{\mu} = 1-\frac{1}{N-1}\sum_{i=1}^{N-1}\frac{\mu_i}{\mu},$$

where $0 \leq B \leq 1$, and

$$\mu = \frac{1}{N}\sum_{i=1}^{N}x_i \qquad \mu_i = \frac{1}{N}\sum_{j=1}^{i}x_j \qquad i = 1, 2, \ldots, N$$

are the general and the partial means for units sorted in non-decreasing order with respect to the X variable.

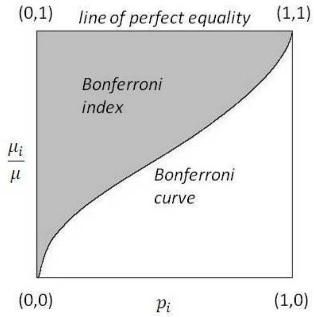
The B index satisfies the axiomatic properties required for inequality indices [i.e. the principle of transfer, of proportional addition to incomes, of proportional addition to persons, of symmetry, of normalization and of operationality] (Giorgi, 1998, p. 142) and $B \geq R$ holds, because it gives bigger weights to units with lower values in the X ranking (see, e.g., De Vergottini, 1950 pp. 318-319 and Pizzetti, 1951 p. 302). Therefore, B is more sensitive to lower levels of the income distribution (see, e.g., Giorgi and Mondani, 1995).

The Bonferroni index is linked to the Bonferroni curve (Figure 1) which is obtained by plotting the cumulative proportion of recipients ($p_i = i/N$), arranged in non-decreasing values of X, versus the corresponding ratio between partial mean and total mean (μ_i/μ).

The polygonal line joining the points (p_i, μ_i/μ) is the Bonferroni curve. If all the recipients in the population have the same quantity of X (i.e equal to μ) the Bonferroni curve coincides with the line of perfect equality that joins the coordinate points (0,0), (0,1), (1,1).

The area between the Bonferroni curve and the line of perfect equality is the concentration area, which is equal to the value of B (Giorgi and Crescenzi, 2001, p. 572-573).

Figure 1 – *An example of Bonferroni curve.*



2. The Shapley decomposition

To overcome the problem of additive decomposition of R by population subgroups, Deutsch and Silber in 2007 used the Shapley decomposition, first introduced in this field by Shorrocks (1999). They derived the impact of four components: inequality within population subgroups (w), inequality between population subgroups (b), ranking (r) and the relative size in each population subgroup (n).

Shapley decomposition is based on the well-known concept of Shapley value in cooperative game theory (Shapley, 1953). The idea of Shapley value is to remove from time to time the contribution of all possible combinations of considered factors for determining their marginal contribution. Therefore, when the method is applied to inequality indices, considering factors in symmetrical manner, it allows to derive the expected marginal contribution to inequality of each factor. Moreover, the contributions sum to the exactly amount of inequality index considered (Shorrocks 1999 and 2013).

For comparing the results obtained by Deutsch and Silber (2007) on R, the same factors have been considered for decomposing B (i.e. w, b, r, n).
Let assume to have a population P partitioned in k population subgroups P_j ($j = 1, \ldots, k$) where y_{ji} is the income of recipient i ($i = 1, \ldots, n_j$) in the population subgroup j.

For removing w is sufficient to replace y_{ji} with μ_j, that is the average income of the population subgroup j to which the recipient i belongs ($y_{ij} \rightarrow \mu_j$). While, for removing b, a kind of standardization is done and y_{ij} must be replaced with $y_{ij}\ \frac{\mu_j}{\mu}$ ($y_{ij} \rightarrow y_{ij}\ \frac{\mu_j}{\mu}$). For closing the effect of difference in size (n) of population subgroups off, the population subgroups must be brought to have the same sizes. Therefore, the least common multiple of size of each population subgroup is done and the y_{ij} are repeated by the number that leads equality in size between the population subgroups. When applying on B, an objection usually raised is that B, on the contrary of R, does not satisfy the Dalton principle of replication invariant. However, with a simulation study – not included here for the sake of brevity – has been verified that, when the size increases (since 100 units), the effect of replications becomes negligible in B. Finally, for removing the effect of ranking (r) it is sufficient to sort, firstly, the population subgroups by their average income, μ_j, and then the recipients by their income within each population subgroup. Obviously for removing the effect of two or more factors at the same time, the methods just illustrated must be applied together.

The marginal impact (SV) of each factor is derived computing the following weighted means of the indices (I=R,B) derived when, from time to time, the effect of components is removed:

$$SV_w = \frac{1}{4}(I - I_w) + \frac{1}{12}[(I_b - I_{wb}) + (I_n - I_{wn}) + (I_r - I_{wr})] + \frac{1}{12}[(I_{bn} - I_{wbn}) + (I_{br} - I_{wbr}) + (I_{rn} - I_{wrn})] + \frac{1}{4}(I_{bnr} - I_{wbnr}) \quad (1)$$

$$SV_b = \frac{1}{4}(I - I_b) + \frac{1}{12}[(I_w - I_{wb}) + (I_n - I_{bn}) + (I_r - I_{br})] + \frac{1}{12}[(I_{wn} - I_{wbn}) + (I_{wr} - I_{wbr}) + (I_{rn} - I_{bnr})] + \frac{1}{4}(I_{wnr} - I_{wbnr}) \quad (2)$$

$$SV_n = \frac{1}{4}(I - I_n) + \frac{1}{12}[(I_w - I_{wn}) + (I_b - I_{bn}) + (I_r - I_{nr})] + \frac{1}{12}[(I_{wb} - I_{wbn}) + (I_{wr} - I_{wnr}) + (I_{br} - I_{bnr})] + \frac{1}{4}(I_{wbr} - I_{wbnr}) \quad (3)$$

$$SV_r = \frac{1}{4}(I - I_r) + \frac{1}{12}[(I_w - I_{wr}) + (I_b - I_{br}) + (I_n - I_{nr})] + \frac{1}{12}[(I_{wb} - I_{wbr}) + (I_{wn} - I_{wnr}) + (I_{bn} - I_{bnr})] + \frac{1}{4}(I_{wbn} - I_{wbnr}) \quad (4)$$

In the expressions (1)-(4) the subscript of I denotes which factor has been removed (for instance I_w is the index computed when the component of within inequality, w, has been removed).

2.1. Numerical illustration

To better explain how the Shapley decomposition works, the example in Deutsch and Silber (2007) is recovered and all the computational steps are exhaustively explained. Let us consider a population with 5 recipients with related income 2, 4, 14, 30 and 50. Assume that individuals with income 2, 14 and 50 belong to population subgroup A and those with income 4 and 30 belong to population subgroup B.

Table 1 shows all the scenarios when removing factors separately, in pairs, in tern and all together. Furthermore, the related income distribution and the values of R and B are presented. Then, the marginal contributions for each factor (SV) have been derived with the expressions (1)-(4) and the values are reported in Table 2.

Because this is an illustrative example on the application of the Shapley decomposition, here the replication invariance principle property for B is overlooked. Anyway, some important preliminary results can be stressed. In particular, when removing w, B is negative (Table 1, case 2 and 7). It can occurs when there is negative correlation between mean income and mean rank (Frick and Goebel, 2008, p. 559). In fact, in the extreme case, when arranging the distribution of income in decreasing order, Rao (1969, p. 245) shows that R is equal to $-R$, and the same occurs for B.

Table 1 – *Gini (R) and Bonferroni (B) indices in different scenarios in which the factors have been removed. Illustrative example related to the income of 5 recipients belonging to two different population subgroups: A={2, 14, 50} and B={4, 30}.*

Removed factor		Income distribution	R	B
1	$-$	2 4 14 30 50	0.488	0.698
2	w	22 17 22 17 22	0.000	-0.017
3	b	1.82 4.71 12.73 35.29 45.45	0.471	0.686
4	n	2 2 4 4 4 14 14 30 30 30 50 50	0.481	0.650
5	r	4 30 2 14 50	0.304	0.431
6	wb	20 20 20 20 20	0.000	0.000
7	wn	22 22 17 17 17 22 22 17 17 17 22 22	0.000	-0.022
8	wr	17 17 22 22 22	0.060	0.098
9	bn	1.82 1.82 4.71 4.71 4.71 12.73 12.73 35.29 35.29 35.29 45.45 45.45	0.462	0.638
10	br	4.71 35.29 1.82 12.73 45.45	0.236	0.346
11	rn	4 4 4 30 30 30 2 2 14 14 50 50	0.284	0.415
12	wbn	19.5 19.5 19.5 19.5 19.5 19.5 19.5 19.5 19.5 19.5 19.5 19.5	0.000	0.000
13	wbr	20 20 20 20 20	0.000	0.000
14	wnr	17 17 17 17 17 17 22 22 22 22 22 22	0.064	0.091
15	bnr	4.71 4.71 4.71 35.29 35.29 1.82 1.82 12.73 12.73 12.73 45.45 45.45 45.45	0.217	0.345
16	$wbnr$	20 20 20 20 20 20 20 20 20 20 20 20	0.000	0.000

Note: w=inequality within, b=inequality between, n=size, r=ranking.

The results in Table 2 provide an initial idea on the hierarchy and the magnitude of the marginal contribution of each factor in determining R and B. Without pursuing further the results of the example, it is just important to point out that the hierarchy of the factors is the same for both the indices and their magnitude change slightly.

Table 2 – *Marginal impact of each component on Gini (R) and Bonferroni (B) indices. Illustrative example related to the income of 5 recipients belonging to two different population subgroup: A={2, 14, 50} and B={4, 30}.*

Factor	Contribution on R		Contribution on B	
	SV	%	SV	%
w	0.353	72.34	0.515	73.77
b	0.038	7.78	0.045	6.39
n	0.005	1.02	0.018	2.64
r	0.092	18.86	0.120	17.20
I	0.488	100.00	0.698	100.00

Note: w=inequality within, b=inequality between, n=size, r=ranking.

3. Application on real data

The Shapley decomposition of Gini ratio index (R) and Bonferroni index (B) have been applied on the data collected by Italian component of European Survey on Income and Living Condition of 2009. The Eu-SILC is a yearly survey carried out in all European countries and defined within the European Regulation no. 1177/2003. Its main aim is to provide data on income, poverty and social exclusion, both cross-sectional and longitudinal. The Italian sample of 2009 survey is 20,928 household and 52,433 individuals. We consider the whole Italian population divided into three population subgroups, that are the main geographical areas: North, Center and South. In table 3, some descriptive statistics on household income distribution for the whole population and for the population subgroups are presented.

The inequality measures have been computed with respect to the household incomes. The incomes have not been equivalised to take into account the different size of the households. The values of R have been estimated through the expression of the sampling estimator defined by Eurostat (2004, p. 39), whilst B through the expression of the sampling estimator derived in Giorgi and Guandalini (2013, p. 154).

Looking at Table 3, North and Center have a quite similar situation. Whilst in South there are lower incomes and higher inequality. Through Shapley decomposition the impact of within inequality (w), between inequality (b), and ranking (r), different size of subgroups (n), both on R and B, has been derived. Then the contribution for each component on these two inequality measures have been compared. The sample size in the three subpopulations considered are larger than 4,000 sampling units, therefore the Dalton principle of replication invariance can be considered satisfied by B, too.

Table 3 – *Some descriptive statistics on average Italian household income distribution by three population subgroups (North, Center and South). Eu-SILC, Italy 2009.*

Total Geografical Area	Percentage of households	Q1	Median	Q3	Mean	R	B
North	48.30	16,603	26,631	40,844	31,872	0.360	0.481
Center	19.74	16,566	26,330	41,066	31,609	0.358	0.481
South	31.96	12,819	20,285	31,272	24,434	0.367	0.488
Italy	100.00	15,148	24,118	38,233	29,442	0.367	0.487

Note: Sample size=20,928; Total number of the household=24,641,200.

In Figure 2 the Lorenz curve and the Bonferroni curve for each scenario have been reported. They show what happen to the income distribution when the components considered are removed separately, in pairs, in tern and all together. In this way it is easier to understand in which way the factors contribute to the inequality.

Table 4 – *Marginal impact of each component on Gini (R) and Bonferroni (B) indices. Confidence interval at 95% in squared brackets. Application on average Italian household income distribution by three population subgroups (North, Center and South). Eu-SILC, Italy 2009.*

Factor	Contribution on R		Contribution on B	
	SV	%	SV	%
w	0.24111 [0.24100, 0.24122]	65.70 [65.69, 65.71]	0.33433 [0.33414, 0.33452]	68.74 [68.71, 68.77]
b	0.03553 [0.03540, 0.03567]	9.68 [9.65, 9.72]	0.04603 [0.04585, 0.04622]	9.46 [9.43, 9.50]
n	0.00131 [0.00128, 0.00134]	0.36 [0.35, 0.36]	0.00544 [0.00534, 0.00554]	1.12 [1.10, 1.14]
r	0.08904 [0.08894, 0.08913]	24.26 [24.24, 24.29]	0.10057 [0.10041, 0.10073]	20.68 [20.65, 20.71]
I	0.36699 [0.36682, 0.36716]	100.00	0.48637 [0.48618, 0.48657]	100.00

Note: w=inequality within, b=inequality between, n=size, r=ranking.

The index R is equal to 0.367 and B is equal to 0.502 (Figure 2: case 1). When removing the component w both the indices are close to 0. Instead, when closing the b component off, they decrease slightly. Removing n has an unimportant effect on indices. Finally, when removing r component, we have a particular trend of both curves, because in these cases we order first the population subgroups and then the units within the population subgroup (Figure 2: cases 5, 10, 11 and 15). In table 4 the absolute and relative impact of each of four components considered is shown.

4. Conclusion and further research

An important topic on inequality measures is their decomposition. Two main lines of research can be distinguished: decomposition by income sources and by population subgroups. In the literature less attention has been paid on decomposition by population subgroups because the inequality indices are not additively decomposable. To overcome this drawback Deutsch and Silber (2007) used the so-called Shapley method to decompose R. In the present paper the same method has been applied to decompose B.

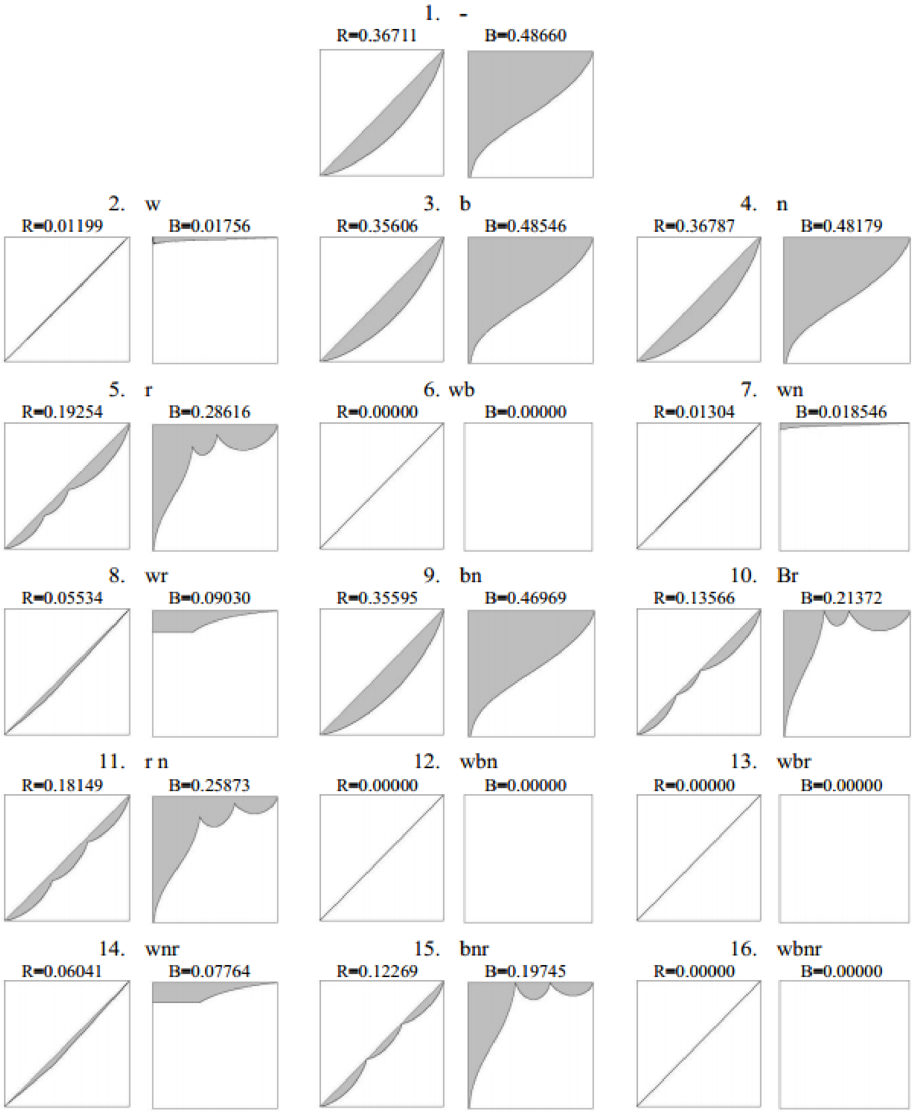
The Shapley decomposition has been useful also to highlight the difference between R and B. The empirical illustration shows the decomposition of both indices in Italy in 2009 when the whole Italian population is divided considering the three main geographical areas: North, Center and South. Four components are considered in the decomposition: inequality within groups (w), inequality between groups (b), differences in size (s) and ranking (r). For both indices, most of the total inequality is due to w followed by r, b and n.

The relative contribution of between groups inequality is similar for both the indices. The within groups inequality has a higher contribution in determining R, whilst ranking and differences in size have a higher contribution in determining B. The study case shows that, besides the difference between B and R in assigning weights to the units and the consequent greater sensitivity for lower levels of income by B, other interesting features make the index different. In fact, the features of each group, such as homogeneity within - denoted by the component of inequality within (w) - and the size of the groups (component n), have higher influence on B than on R. The hierarchy and the magnitude of these components in determining the inequality appear to be confirmed both in the application on real data and in the numerical illustration. However, a deeper investigation on the range of variation of the components under different income distributions is very interesting and it will be the object of further studies.

Acknowledgment

The present work has been realized within the grant for the project “Indici di disuguaglianza e variabilità: nuove prospettive di ricerca” (Sapienza 2013).

Figure 2 – *Gini curve and Bonferroni curve when removing the component within (w), between (b), size (n) and ranking (r) separately, in pairs, in tern and all together.*



References

BÁRCENA-MARTIN E., SILBER J. 2013. On the generalization and decomposition of the Bonferroni index. *Social Choice and Welfare*, Vol. 41, No. 4, pp. 763-787.

BONFERRONI C.E. 1930. *Elementi di statistica generale*. Firenze: Libreria Seber.

DEUTSCH J., SILBER J. 2007. Decomposing income inequality by population subgroups: a generalization. In J.A. Bishop and Y. Amiel (eds.), *Research on Economic Inequality: Inequality and Poverty*. Berlin: Springer, Vol. 14, pp. 237-253.

DE VERGOTTINI M. 1950. Sugli indici di concentrazione. *Statistica*, Vol. 10, pp. 445-454.

EUROSTAT 2004. Common cross-sectional EU indicators based on EU-SILC; the gender gap. *Working Group on Statistics on Income and Living Conditions (EU-SILC)*, 29-30 March, Luxemburg, pp. 1-42.

FRICK J.R., GOEBEL J. 2008. Regional Income Stratification in Unified Germany Using a Gini Decomposition Approach. *Regional Studies*, Vol. 42, No. 4, pp. 555-577.

GINI C. 1914. Sulla misura della concentrazione e della variabilità dei caratteri, *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Vol. 73, pp. 1203-1248. (English translation in *Metron*, 2005, Vol. 63, pp. 3-38).

GIORGI G.M. 1998. Concentration index, Bonferroni. In Kotz S. *et. al* (Eds.) *Encyclopedia of Statistical Sciences*, Update 2. New York: Wiley-Intersciences, pp. 141-146.

GIORGI G.M. 2011. The Gini inequality index decomposition, an evolutionary study. In Deutsch J., Silber J. (Eds.) *The measurement of individual well-being and group inequality: essay in memory of Z.M. Berrebi*. London: Routledge, pp. 185-218.

GIORGI G.M., CRESCENZI M. 2001. A look at the Bonferroni inequality measure in a reliability framework. *Statistica*, Vol. 61, No. 4, pp. 571 -583.

GIORGI G.M., GUANDALINI A. 2013. A sampling estimator of the Bonferroni inequality index. *Rivista Italiana di Economia, Demografia e Statistica*, Vol. LXVII, No. 3/4, pp. 151-158.

GIORGI G.M., MONDANI R. 1995. Sampling distribution of Bonferroni inequality index from an exponential population. *Sankhya*, Series B, Vol. 57, No. 1, pp. 10-18.

NYGÅRD F., SANSTRÖM A. 1981. *Measuring income inequality*. Stockholm: Almqvist & Wiksell International.

PIESCH W. 1975. *Statistische Konzentrationsmasse*. Tübingen: J.B.C. Mohr (Paul Siebeck).

PIZZETTI E. 1951. Relazioni fra indici di concentrazione. *Statistica*, Anno XI, No. 3-4, pp. 294-316.

RAO V. 1969. Two decompositions of concentration ratio. *Journal of the Royal Statistical Society*, Vol 132A, pp. 418-425.

SHAPLEY L. 1953. A value for n-person games. In Kuhn H.K. and Tucker A.W. (Eds.) *Contributions to the theory of games, 2*. Princeton (N.J.): Princeton University Press, pp. 307-315.

SHORROCKS A.F.1999. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value: *University of Essex – Department of Economics, Unpublished Paper*.

SHORROCKS A.F. 2013. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *The Journal of Economic Inequality*, Vol. 11, No. 1, pp. 99-126.

TARSITANO A. 1990. The Bonferroni index of income inequality. In Dagum C. and Zenga M. (Eds.) *Income and Wealth Distribution, Inequality and Poverty*. Berlin: Springer-Verlag, pp. 228-242.

SUMMARY

Bonferroni Index Decomposition and the Shapley method

The Bonferroni inequality index (B) remained almost forgotten until the last two decades. Recently, it has been rediscovered and furthermore, new and interesting interpretations of B have been proposed. An important topic in the literature on inequality measures is their decomposition. Two main important lines of research involve decomposition by income sources and by population subgroups. Many contributions are related to R, less to B. The Shapley decomposition enables to overcome the problem related to inequality index of not being additively decomposable into the sum of within and between groups components. In this perspective Deutsch and Silber (2007) use the Shapley decomposition for R. In this paper, the Shapley decomposition have been applied to B, too. The comparison among the results obtained for both the indices allows to highlight other interesting similarities and differences among the two indices.

Giovanni Maria GIORGI, Department of Statistical Sciences, "Sapienza" University of Rome, giovannimgiorgi@gmail.com

Alessio GUANDALINI, Italian National Institute of Statistics – ISTAT, alessio.guandalini@istat.it