zef
Center for
Development Research
University of Bonn

ZEF-Discussion Papers on

Development Policy No. 343

Oded Stark and Wiktor Budzinski

The merger of populations as a revision of comparison space: Repercussions for social stress and income inequality

Bonn, April 2024.

Oded Stark and Wiktor Budzinski. The merger of populations as a revision of comparison space: Repercussions for social stress and income inequality. ZEF Discussion Papers on Development Policy No. 343, Center for Development Research, Bonn, April 2024, pp. 14.

ISSN: 1436-9931

Published by:
Zentrum für Entwicklungsforschung (ZEF)
Center for Development Research
Genscherallee 3
D – 53113 Bonn
Germany
Phone: +49-228-73-1861
Fax: +49-228-73-1869
E-Mail: zef@uni-bonn.de
www.zef.de

The authors:

Oded Stark, University of Bonn and University of Warsaw. Contact: ostark@uni-bonn.de

Wiktor Budzinski, University of Warsaw.

Acknowledgements

We are indebted to a reviewer for a thoughtful commentary, and to Audra Bowlus for advice and guidance.

Abstract

The merger of populations expands the comparison space of incomes. As a result, measures of the income-based social stress and of the income inequality of the constituent populations need to be replaced by new measures. To this end, we develop a procedure for calculating the aggregate social stress and the Gini coefficient of the merged population. We show that to calculate the aggregate social stress when the income distributions of the constituent populations do not overlap, it is sufficient to utilize just three characteristics of the constituent populations: their size, the levels of their aggregate income-based social stress, and their mean income. This result carries over to the calculation of the Gini coefficient of the merged population. We also analyze the extent to which the procedure, applied to cases where the constituent populations do not overlap, can be extended to cases where the income distributions of the constituent populations overlap.

Keywords: Gini coefficient; Stress of a population; Merger of populations; Income inequality of a merged population

JEL classification: D31; D63; F02; F15

1. Introduction

Mergers of populations occur often, and in many spheres: they may arise naturally or as a result of administrative considerations, and they may be imposed or voluntarily accepted. Provinces consolidate into regions, small municipalities merge into larger metropolitan areas, adjacent villages experiencing population growth coalesce into towns, and so on. Governments merge administrative units because of a presumption that doing so will reduce duplication and costly outlays, streamline bureaucracies, and increase efficiency and productivity brought about by scale economies, for example. After all, classical trade theory maintains that integration liberalizes trade and smoothes labor and financial flows, and that larger and denser markets improve resource allocation and the distribution of final products. Consequently, it is posited, the welfare of the integrating populations will rise.

When populations merge, individuals' comparison space of incomes expands, and the set of comparators of some individuals changes. As a result, measures of income-based social stress and of income inequality of the constituent populations need to be replaced by new measures. Quite often, and perhaps more often than not, in population profiles, a standard feature used to characterize a population is the Gini coefficient of the distribution of the population's income. So, when populations merge, it is natural to inquire about the Gini coefficient of the distribution of income of the merged population. It is also natural to ponder whether there is a need to calculate this coefficient from scratch or, alternatively, whether the coefficient could be gleaned from data on the Gini coefficients of the constituent populations when they were separate entities. In this paper, we develop an approach to doing the latter.

We pursue a three-step procedure. First, following Sen (1973 and 1997), we express the Gini coefficient of a population as the population's aggregate income-based social stress divided by the population's aggregate income. Second, we develop a formula for calculating the aggregate income-based social stress of a population resulting from the merger of two populations. The formula pertains to the case in which the incomes of the constituent populations do not overlap. The formula is lean in requirements: all that is needed is information on the number of members of each of the constituent populations, the levels of aggregate stress of the constituent populations, and the mean incomes of the constituent populations. Third, combining the formula with information on the aggregate incomes of the constituent populations and, hence, on the aggregate income of the merged population leads us directly to the Gini coefficient of the merged population; calculation from scratch of the

Gini coefficient of the merged population is not required. Lastly, we replicate the preceding steps for the case in which the incomes of the constituent populations overlap.

2. Calibrating income-based social stress

In population $N = \{1, 2, \ldots, n\}$, $n \geq 2$, let $y = (y_1, ..., y_n)$ be the vector of incomes of the members of the population. Let these incomes be ordered: $0 < y_1 < y_2 < \ldots < y_n$. RD_i - by which we denote the income-related social stress of individual *i*, $i = 1, 2, ..., n-1$, whose income is y_i - is defined as

$$RD_i \equiv \frac{1}{n} \sum_{j=i+1}^{n} (y_j - y_i) , \tag{1}$$

where it is understood that $RD_n \equiv 0$.

The idea here is to aggregate the income excesses (the differences between the incomes that are higher than the income of individual *i* and the income of individual *i*) and normalize this sum, that is, divide it by the size of the population. Because the stress of an individual stems from having an income that is lower than the incomes of others (rather than from having a low absolute income), we refer to this stress as income-based social stress or as income-based relative deprivation. A detailed derivation of this representation of an individual's social stress is provided at the end of this paper in Appendix: Construction of a relative deprivation index.[1]

We denote by *TRD* the sum (aggregate) of the levels of RD_i in population *N*:

$$TRD \equiv \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (y_j - y_i) \ . \tag{2}$$

[1] We characterize the stress that arises from having less than others as social, and we quantify this stress by (1). In taking this approach we follow, and are aligned with, a large body of literature on the subject of relative deprivation and reference (comparison) groups, spanning from the pioneering 1949 two-volume study of Stouffer et al. (1949a, 1949b), through Akerlof (1997), and all the way to recent writings, for example of Stark et al. (2017) and Stark (2020). The latter two studies include deliberations and discussions on the identity of the reference group, and they provide many references to related works. By definition and construction, relative deprivation is the dual of the concept of reference group or comparison group, hence the term social.

3. The Gini coefficient

Following Sen (1973), the Gini coefficient, *G*, of population $N=\{1,2,\ldots,n\}$, $n\geq 2$, with a vector $y=(y_1,...,y_n)$ of the incomes of the members of the population, is defined as

$$G\equiv\frac{\sum_{j=1}^{n}\sum_{i=1}^{n}\left|y_i-y_j\right|}{2n^2\bar{y}}, \tag{3}$$

where $\bar{y}=(1/n)\sum_{i=1}^{n}y_i$ is the average income of the population. In Sen's (1973, p. 8) words: "In any pair-wise comparison the man with the lower income can be thought to be suffering from some depression on finding his income to be lower. Let this depression be proportional to the difference in income. The sum total of all such depressions in all possible pair-wise comparisons takes us to the Gini coefficient." In this paper we use income-based "depression" and income-based stress interchangeably.

On noting that $\sum_{j=1}^{n}\sum_{i=1}^{n}\left|y_i-y_j\right|=2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(y_j-y_i)$, an equivalent representation of the Gini coefficient in (3), which disposes of the need to operate with absolute values, is

$$G=\frac{\frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(y_j-y_i)}{\sum_{i=1}^{n}y_i}=\frac{TRD}{TI}. \tag{4}$$

Thus, the Gini coefficient in (4) is expressed as a ratio: *TRD* as defined in (2), divided by aggregate (total) income $\sum_{i=1}^{n}y_i=TI$.

4. A procedure for calculating the *TRD* and the Gini coefficient of a merged population

4.1 The case of nonoverlapping populations

Let there be two populations, *M* and *N*, and let there be m individuals in population *M*, and n individuals in population *N*. We denote the aggregate or the total income-based stress of each of these two populations when apart by TRD_M and TRD_N, respectively. Let the incomes of the individuals in *M* be $x_1<x_2<...<x_m$, the incomes of the individuals in *N* be $y_1<y_2<...<y_n$, and the highest income in population *M* be lower than the lowest income in

population *N*, that is, let $x_m < y_1$. Thus, population *M* is relatively poor, and population *N* is relatively rich. We denote the mean incomes of populations *M* and *N* by μ_M and μ_N, respectively. Obviously, $\mu_M < \mu_N$. We denote by $TRD_{M\cup N}$ the total income-based stress of the population formed by the merger of populations *M* and *N*. We have the following claim.

Claim 1. $TRD_{M\cup N} = \frac{1}{m+n}[mTRD_M + nTRD_N + mn(\mu_N - \mu_M)]$.

Proof. In the Appendix.

We denote the Gini coefficient of population $M \cup N$ by $G_{M\cup N}$, and the aggregate income of population $M \cup N$ by $TI_{M\cup N}$. Thus, $TI_{M\cup N} = m\mu_M + n\mu_N$. We have the following remark.

Remark 1. $G_{M\cup N} = \frac{TRD_{M\cup N}}{TI_{M\cup N}}$.

The remark follows directly from (4) and Claim 1.

Because when we calculate both $TRD_{M\cup N}$ and $TI_{M\cup N}$ all that is needed is information on the number of members of each of the constituent populations, the levels of the aggregate stress of the constituent populations, and the mean incomes of the constituent populations, we can use these measurements to calculate the Gini coefficient of the merged population. A direct calculation is not required.

Example 1. Let $x_1 = 1,\ x_2 = 2,\ y_1 = 3, y_2 = 4.$ A direct calculation of the Gini coefficient of this population, as per (4), yields $\frac{\frac{1}{4}(1+2+3)+\frac{1}{4}(1+2)+\frac{1}{4}\cdot 1}{10} = \frac{1}{4}.$ A calculation based on Claim 1 and Remark 1 yields $\frac{\frac{1}{4}(2\cdot\frac{1}{2}+2\cdot\frac{1}{2}+2\cdot 2\cdot 2)}{10} = \frac{1}{4}.$

4.2 The case of overlapping populations

We next relax the assumption that the two populations do not overlap. As before, we have population *M* of *m* individuals, population *N* of *n* individuals, and that the income distributions of the two populations are given, respectively, by $x_1 < x_2 < ... < x_m$ and $y_1 < y_2 < ... < y_n$. However, we now allow for the possibility that the highest income in population *M*, x_m, is higher than the lowest income in population *N*, y_1. We refer to such a circumstance as an overlap. We have the following claim.

Claim 2. $TRD_{M\cup N} = \dfrac{1}{m+n}\left[mTRD_M + nTRD_N + \sum_{i=1}^{m}\sum_{k=1}^{n} | x_i - y_k | \right].$

Proof. In the Appendix.

Remark 2. $G_{M\cup N} = \dfrac{TRD_{M\cup N}}{TI_{M\cup N}}$.

The remark follows directly from (4) and Claim 2.

Example 2. Let $x_1 = 1,\ x_2 = 3,\ y_1 = 2, y_2 = 4.$ As in Example 1, a direct calculation of the Gini coefficient of this population, as per (4), yields $\dfrac{\frac{1}{4}(1+2+3)+\frac{1}{4}(1+2)+\frac{1}{4}\cdot 1}{10} = \dfrac{1}{4}$. A calculation based on Claim 2 and Remark 2 yields $\dfrac{\frac{1}{4}(2\cdot 1+2\cdot 1)+\frac{1}{4}\cdot 6}{10} = \dfrac{1}{4}$.

Example 3. Let $x_1 = 1,\ x_2 = 4,\ y_1 = 2, y_2 = 3.$ As in Examples 1 and 2, a direct calculation of the Gini coefficient of this population, as per (4), yields $\dfrac{\frac{1}{4}(1+2+3)+\frac{1}{4}(1+2)+\frac{1}{4}\cdot 1}{10} = \dfrac{1}{4}$. A calculation based on Claim 2 yields $\dfrac{\frac{1}{4}(2\cdot\frac{3}{2}+2\cdot\frac{1}{2})+\frac{1}{4}\cdot 6}{10} = \dfrac{1}{4}$.

5. Implications

From Claim 1 we see that the total income-based stress of a population formed by the merger of populations is equal to a weighted sum of the levels of the total income-based stress of the constituent populations - where the weights are the shares of the constituent populations in the merged population - plus “a residual.” This representation is interesting because a high TRD_M will not measurably affect $TRD_{M\cup N}$ if the weight $\dfrac{m}{m+n}$ is small, and, similarly, a high TRD_N will not measurably affect $TRD_{M\cup N}$ if the weight $\dfrac{n}{m+n}$ is small. The “residual” is a measure of closeness: when the mean incomes of the constituent populations are similar, the residual is small, and when these mean incomes differ significantly, the residual is large.

From $G_{M\cup N}=\frac{TRD_{M\cup N}}{TI_{M\cup N}}$ and Claim 1 we can see how the Gini coefficient of the merged population can be decomposed by source or, putting it differently, how the income inequalities of the constituent populations contribute to the income inequality of the merged population. In particular, given the mean income of the richer population, the higher the mean income of the poorer population is, the lower $G_{M\cup N}$ is. What is also interesting is that it is not the case that one of the two constituent populations will exert a greater influence on income inequality because of the magnitude of its *TRD*. Either of the two constituent populations, if its *TRD* is high, will have less of an impact on the Gini coefficient of the merged population if its share in the combined population is small. The impact of the richer population on $G_{M\cup N}$ is mitigated by its TRD_N if that TRD_N is small, and by its relative size when that size is small. While both the absolute size and the relative size of a constituent population matter, it is not the case that the absolute or relative size of one population will inherently matter more than the absolute or relative size of the other population.

Common to Examples 1, 2, and 3 is that each of them is of a population of two individuals. In such a setting, the Gini coefficient of the merged population will never be smaller than *both* the Gini coefficients of the constituent populations. To see this, we consider the case in which the incomes of population *M* are $0<x_1<x_2$, the incomes of population *N* are $0<y_1<y_2$, and $x_1<y_1<x_2<y_2$.[2] In this case, $G_M=\frac{x_2-x_1}{2(x_1+x_2)}$, $G_N=\frac{y_2-y_1}{2(y_1+y_2)}$, and $G_{M\cup N}=\frac{-3x_1-y_1+x_2+3y_2}{4(x_1+y_1+x_2+y_2)}$. If the Gini coefficients of populations *M* and *N* were both larger than the Gini coefficient of the merged population, then the following inequalities would have to hold:

[2] This observation can easily be generalized to the case of the merger of any two populations of two individuals each, as long as the incomes of these populations are not identical.

$$\begin{cases} \dfrac{x_2 - x_1}{2(x_1 + x_2)} > \dfrac{-3x_1 - y_1 + x_2 + 3y_2}{4(x_1 + y_1 + x_2 + y_2)} \\ \dfrac{y_2 - y_1}{2(y_1 + y_2)} > \dfrac{-3x_1 - y_1 + x_2 + 3y_2}{4(x_1 + y_1 + x_2 + y_2)}. \end{cases}$$

Upon rearrangement, these inequalities simplify to

$$\begin{cases} (x_1 + x_2)^2 > y_2(x_2 + 5x_1) - y_1(3x_2 - x_1) \\ (y_1 + y_2)^2 < y_2(x_2 + 5x_1) - y_1(3x_2 - x_1). \end{cases} \tag{5}$$

Because the right-hand sides of the two inequalities in (5) are the same, this implies that $(x_1 + x_2)^2 > (y_1 + y_2)^2$ which, given that $x_1 < y_1$ and that $x_2 < y_2$, is a contradiction. Thus, it cannot be the case that the Gini coefficient of the merged population will be smaller than *both* the Gini coefficients of the constituent populations. An implication of this finding is that, in and of itself, the merger of one population with another population cannot serve as a policy tool for reducing the inequality of both populations.

The merger of populations may involve $n > 2$ constituent populations. For such a case, we do not need a distinct protocol for calculating the Gini coefficient. In the spirit of a proof by induction, the reason is that the *n* populations can be merged sequentially; that is, we first merge two populations and then merge this new population with a third population; we then merge this new population with a fourth population, and so on. In doing this, we draw on formulas that are readily available in Claims 1 and 2 and in Remarks 1 and 2. A second, and probably somewhat more efficient, approach is to use the actual induction protocol to obtain a specific formula for the case of $n > 2$. This is a rather simple, albeit rather involved, algebraic exercise. To save space, the exposition is omitted here, but available upon request.

6. Conclusion

Two appealing advantages of calculating the Gini coefficient of a merged population by utilizing merely the sizes, the levels of the aggregate income-based social stress, and the mean incomes of the constituent populations, are that this procedure affords an insight into the contribution to inequality by source, and it delivers an efficiency gain: we can identify the roles played by the different factors that "feed" into the coefficient, that is, "who contributes what and by how much," and we can "get there" quite easily. For example, if there are 10 incomes in each of two constituent populations, then in order to obtain the Gini coefficient of the merged population, there is no need, as per (3), to calculate 400 income comparisons; instead, by Claim 1 and Remark 1, we can obtain the coefficient by the mere few calculations of *TRD* and *TI*.

Appendix: **Proofs of Claims 1 and 2**

Proof of Claim 1. From the assumption that $x_m < y_1$, we know that the individuals from *N* do not experience income-based stress from having incomes that are lower than the incomes of the individuals in *M*. Using this fact and the definition of *TRD* in (2), we know that unfavorable income comparisons occur in three parts:

$$TRD_{M\cup N} = \frac{1}{m+n}\left[\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}(x_j - x_i) + \sum_{k=1}^{n-1}\sum_{l=k+1}^{n}(y_l - y_k) + \sum_{i=1}^{m}\sum_{k=1}^{n}(y_k - x_i)\right]. \qquad \text{(A1)}$$

The first two double sums in (A1) are clearly $mTRD_M$ and $nTRD_N$, respectively, whereas the third double sum in (A1) is the contribution to the *TRD* of the merged population that arises from the comparisons of the incomes of the members of the poorer population *M* with the incomes of the members of the richer population *N*. This third double sum can be developed as follows.

$$\begin{aligned}\sum_{i=1}^{m}\sum_{k=1}^{n}(y_k - x_i) &= \sum_{i=1}^{m}(\sum_{k=1}^{n} y_k - nx_i) = n\sum_{i=1}^{m}(\mu_N - x_i)\\ &= n(m\mu_N - \sum_{i=1}^{m} x_i) = mn(\mu_N - \mu_M)\,.\end{aligned} \qquad \text{(A2)}$$

Then, upon replacing $\sum_{i=1}^{m-1}\sum_{j=i+1}^{m}(x_j - x_i)$ with $mTRD_M$, replacing $\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}(y_l - y_k)$ with $nTRD_N$, and inserting the last part of (A2) into (A1), we obtain

$$TRD_{M\cup N} = \frac{1}{m+n}[mTRD_M + nTRD_N + mn(\mu_N - \mu_M)]\,. \qquad \text{(A3)}$$

Q.E.D.

Proof of Claim 2. To prove the claim, we first rewrite TRD in a form that allows for a more convenient mathematical treatment.

Lemma 1. Let a population M of m individuals with incomes $x_1 < x_2 < ... < x_m$ be given. Then

$$TRD_M = \frac{1}{2m}\sum_{k=1}^{m}\sum_{i=1}^{m}|x_k - x_i|\,. \qquad \text{(A4)}$$

Proof of the Lemma. For all $i,k = 1,\dots,m$, $i \neq k$, either $x_k - x_i > 0$ or $x_i - x_k > 0$. *TRD* in (2) includes only nonnegative differences between incomes in a distribution. Because the *TRD* expression in (A4) includes the absolute values of *all* the differences between incomes, it counts a difference between a pair of given incomes twice. Thus, we can write

$$\frac{1}{m}\sum_{i=1}^{m}\sum_{k=1}^{m}\left|x_k - x_i\right| = 2\frac{1}{m}\sum_{i=1}^{m-1}\sum_{k=i+1}^{m}(x_k - x_i)\,, \tag{A5}$$

or

$$\frac{1}{2m}\sum_{i=1}^{m}\sum_{k=1}^{m}\left|x_k - x_i\right| = \frac{1}{m}\sum_{i=1}^{m-1}\sum_{k=i+1}^{m}(x_k - x_i)\,. \tag{A6}$$

Because the right-hand side of (A6) is (2), a replacement of y_i with x_i notwithstanding, the left-hand side of (A6) is an alternative expression of TRD_M .

Q.E.D.

We now use Lemma 1 to prove Claim 2. We consider how TRD "behaves" upon the merger of two populations that overlap. Using (A4), we obtain

$$TRD_{M\cup N} = \frac{1}{2(n+m)}\left[\sum_{i=1}^{m}\sum_{j=1}^{m}\left|x_j - x_i\right| + \sum_{k=1}^{n}\sum_{l=1}^{n}\left|y_l - y_k\right| + 2\sum_{i=1}^{m}\sum_{k=1}^{n}\left|x_i - y_k\right|\right]. \tag{A7}$$

The first two double sums in (A7) are clearly $2mTRD_M$ and $2nTRD_N$, respectively. We therefore have that

$$TRD_{M\cup N} = \frac{1}{m+n}\left[mTRD_M + nTRD_N\right] + \frac{1}{m+n}\sum_{i=1}^{m}\sum_{k=1}^{n}\left|x_i - y_k\right|. \tag{A8}$$

Q.E.D.

References

Akerlof, George A. (1997). "Social distance and social decisions." *Econometrica* 65: 1005-1027.

Sen, Amartya K. (1973 and 1997). *On Economic Inequality*. Oxford: Clarendon Press.

Stark, Oded (2020). "Relative deprivation as a cause of risky behaviors." *Journal of Mathematical Sociology* 44(3): 138-146.

Stark, Oded, Bielawski, Jakub, and Falniowski, Fryderyk (2017). "A class of proximity-sensitive measures of relative deprivation." *Economics Letters* 160: 105-110.

Stouffer, Samuel A., Suchman, Edward A., DeVinney, Leland C., Star, Shirley A., and Williams Jr., Robin M. (1949a). *The American Soldier*: *Adjustment During Army Life*, Vol. I. *Studies in Social Psychology in World War II*. Princeton: Princeton University Press.

Stouffer, Samuel A., Lumsdaine, Arthur A., Lumsdaine, Marion H., Williams Jr., Robin M., Smith, Brewster M., Janis, Irving L., Star, Shirley A., and Cottrell Jr., Leonard S. (1949b). *The American Soldier*: *Combat and its Aftermath*, Vol. II. *Studies in Social Psychology in World War II*. Princeton: Princeton University Press.

Appendix: Construction of a relative deprivation index

For the purpose of constructing a measure of relative deprivation, a natural starting point is the work of Runciman (1966), who argues that an individual has an unpleasant sense of being relatively deprived when he lacks a desired good and perceives that others with whom he naturally compares himself possess that good. Runciman (1966, p. 19) writes as follows: "The more people a man sees promoted when he is not promoted himself, the more people he may compare himself with in a situation where the comparison will make him feel deprived," implying that the deprivation from not having, say, income y is an increasing function of the fraction of people in the individual's reference group who have y. To aid intuition, we resort to income-based comparisons, that is, an individual feels relatively deprived when others in his reference group earn more than he does. It is assumed here implicitly that the earnings of others are publicly known. Alternatively, we can consider consumption, which might be more publicly visible than income, although these two variables can reasonably be assumed to be strongly positively correlated.

As an illustration of the relationship between the fraction of people possessing income y and the deprivation of an individual lacking y, consider a population (reference group) of six individuals with incomes $\{1,2,6,6,6,8\}$. Imagine a furniture store that in three separate departments sells chairs, armchairs, and sofas. An income of 2 allows you to buy a chair. To buy an armchair, you need an income that is slightly higher than 2. To buy a sofa, you need an income that is slightly higher than 6. Thus, when you go to the store and your income is 2, what are you "deprived" of? Armchairs and sofas. Mathematically, this deprivation can be represented by $P(Y>2)(6-2)+P(Y>6)(8-6)$, where $P(Y>y_i)$ stands for the fraction of those in the population whose income is higher than y_i, for $y_i = 2,6$. The reason for this representation is that when you have an income of 2, you cannot afford anything in the department that sells armchairs, and you cannot afford anything in the department that sells sofas. Because not all those who are to your right in the income distribution sorted in ascending order can afford to buy a sofa, yet they can all afford to buy armchairs, a breakdown into the two (weighted) terms $P(Y>2)(6-2)$ and $P(Y>6)(8-6)$ is needed. In this way, we already get to the essence of the measure of relative deprivation: we take into account the fraction of the reference group (population) of individuals who possess some good that you do not, and we weigh this fraction by the "excess value" of that good. Because income enables an individual to afford the consumption of certain goods, we refer to comparisons based on income.

Formally, let $y=(y_1,...,y_m)$ be the vector of incomes in population *N* of size *n* with relative incidences $p(y)=\left(p(y_1),...,p(y_m)\right)$, where $m\le n$ is the number of distinct income levels in *y*, and *n* and *m* are natural numbers. The relative deprivation, *RD*, of an individual earning y_i is defined as the weighted sum of the excesses of incomes higher than y_i such that each excess is weighted by its relative incidence, that is,

$$RD_N(y_i)\equiv\sum_{y_k>y_i}p(y_k)(y_k-y_i).$$

In the previously given example with income distribution {1,2,6,6,6,8}, the vector of incomes was $y=(1,2,6,8)$ and the corresponding relative incidences were $p(y)=(1/6,1/6,3/6,1/6)$. Therefore, the *RD* of the individual earning 2 was $\sum_{y_k>y_i}p(y_k)(y_k-y_i)=p(6)(6-2)+p(8)(8-2)=\frac{3}{6}\cdot4+\frac{1}{6}\cdot6=3$. By similar calculations, here we see that the *RD* of the individual earning 1 is higher at $3\frac{5}{6}$, and that the *RD* of each of the individuals earning 6 is lower at $\frac{1}{3}$.

We expand the vector y to include incomes with their possible respective repetitions, that is, we include each y_i as many times as its incidence dictates, and we assume that the incomes are ordered, that is, $y=(y_1,...,y_n)$ such that $y_1\le y_2\le...\le y_n$. In this case, the relative incidence of each y_i, $p(y_i)$, is $1/n$, and thus, for $i=1,...,n-1$, we obtain

$$RD_N(y_i)\equiv\frac{1}{n}\sum_{k=i+1}^{n}\left(y_k-y_i\right).$$

This formula is analogous to (1) for RD_i presented in the main body of the paper.

Reference

Runciman, Walter G. (1966). *Relative Deprivation and Social Justice*. Berkeley: University of California Press.