University of Warsaw
Faculty of Economic Sciences

Working Papers

No. 24/2011 (64)

Martyna Kobus
Piotr Miłoś

Inequality decomposition by population subgroups for ordinal data

Warsaw 2011

Inequality decomposition by population subgroups for ordinal data

Martyna Kobus
University of Warsaw
Faculty of Economic Sciences
e-mail: mkobus@wne.uw.edu.pl

Piotr Miłoś
University of Warsaw
Faculty of Mathematics, Informatics and Mechanics
e-mail: pmilos@mimuw.edu.pl

Abstract

We present a class of decomposable inequality indices for ordinal data (e.g. self-reported health survey). It is characterized by well-known inequality axioms (e.g. scale invariance) and a decomposability axiom which states that an index can be represented as a function of inequality values in subgroups and subgroup sizes. The only decomposable indices are strictly monotonic transformations of the weighted average of frequencies in categories. Among the indices proposed in the literature only the absolute value index (Abul Naga and Yalcin, 2008; Apouey, 2007) is decomposable. As an empirical illustration we calculate regional contributions to overall health inequality in Switzerland.

Keywords:
ordered response health data; inequality measurement; health inequality; ordinal data; decomposition

JEL:
I3, I1

Acknowledgements:
Acknowledgements: We would like to thank to Ramses H. Abul Naga for introducing us into the topic of inequality measurement for ordinal data and for helpful advice.

Introduction

In measuring inequality we are typically interested in the dispersion of inequality values with respect to sex, race, gender, age, education level and other characteristics. Disaggregated analysis of the causes of inequality is of tremendous importance for policy because it enables policy makers to target particular inequalities most effectively. Therefore decomposition by population subgroups is considered a highly desired property of inequality indices. This paper presents the class of decomposable indices for ordinal data.[1]

Health surveys in which individuals are asked to choose their health status out of several options are an example of ordinal data that are frequently used in policy and theoretical analyses. In fact, many well-being dimensions, along which inequality can be measured, are of qualitative nature (e.g. happiness, educational attainment). With this type of variables numerical values are assigned to each option and they constitute a *scale*. A distinctive feature of qualitative data is that order is the only relevant information. Formally, since increasing transformations of a given scale all reflect the same ordering of categories, it does not matter which particular transformation is chosen; they are all equivalent. Therefore an index should be invariant to rescalings of variables which preserve the order of categories. It is well-known that conventional inequality indices do not have this property (Zheng, 2011; Allison and Foster, 2004). The formulas of the Gini coefficient, the Atkinson index, the Theil index all depend on the mean which is sensitive to rescalings. An example will clarify.

Suppose the distributions of self-reported health status among men and women are, respectively, $\pi = (0.2, 0.2, 0.2, 0.2, 0.2)$ and $\omega = (0.3, 0.2, 0.1, 0.1, 0.3)$. That is, there are twenty percent men in each health category, thirty percent women in the first category etc. By assumption, higher category number indicates better health status. We consider two scales: $c = (1, 2, 3, 4, 5)$ and $\tilde{c} = (1, 2, 3, 4, 100)$; note that both correspond to the same order of health categories. Then, under scale c the

[1] We use the following names interchangeably: ordinal data, ordered response data, qualitative data.

Gini index for the men's distribution is $GINI(\pi, c) = 0.26$ whereas for women's distribution we get $GINI(\omega, c) = 0.31$, hence health inequality is lower among men than women.[2] However, under scale $\tilde{c}$ the ranking is reversed; $GINI(\pi, \tilde{c}) = 0.72 > GINI(\omega, \tilde{c}) = 0.66$. Clearly, conventional inequality measures are not well-suited for ordinal data. Accordingly, unidimensional indices for ordered response data were introduced by Blair and Lacy (2000), Allison and Foster (2004), Abul Naga and Yalcin (2008), Apouey (2007) and Zheng (2010). Yet these authors did not study decomposability. This paper fills this void.

We characterize decomposable indices in terms of standard inequality axioms (e.g. scale invariance and normalization).[3] The axiom that is similar in spirit to Pigou-Dalton Transfer axiom is called EQUAL and was defined by Allison and Foster (2004). They postulate that a cumulative distribution P reflects more inequality than a cumulative distribution Q if P is obtained from Q via a sequence of median preserving spreads. The intuition is that P is less concentrated around the median than Q.

Following the classic article in the study of inequality decomposition by Shorrocks (1984), we call an index decomposable if it can be represented as some function of subgroups inequality values and sizes.[4] In addition, an index potentially depends on the scale; unless it fulfills scale independence, which is one of the postulated axioms. As the main result (Theorem 3) we provide the functional form for the indices which are decomposable and fulfill scale independence, normalization, continuity and EQUAL. These indices belong to the class of continuous and strictly increasing (decreasing) transformations of an index which is the weighted average of frequencies in particular categories. In addition, weights increase (decrease)

[2]We calculated the Gini index by assuming there are two men in each health category, three women in the first health category, two women in the second health category and the like. This is valid since the Gini index is replication invariant.

[3]Appropriate modifications to account for the ordinal nature of the data were proposed by Abul Naga and Yalcin (2008).

[4]To be precise, in the definition of the decomposability in the sense of Shorrocks (1984) the function can also depend on subgroups means, however, as already noted such requirement would not make much sense in the current setting.

with the distance from the median category. Replacing scale independence with a weaker requirement which is scale invariance (the ordering of the distributions induced by an index is invariant with respect to scale, but the value of an index depends on the scale) adds only the dependence on the scale (Theorem 2). If we do not require that EQUAL holds, then the class of decomposable indices is of course wider but not significantly. That is, decomposable indices are continuous and strictly monotonic transformations of the weighted average of frequencies (Theorem 1). Similarly to conventional indices, where only monotonic transformations of Generalized Entropy indices are decomposable (Shorrocks, 1984), decomposability turns out to be particularly effective in filtering out inequality measures for qualitative data. Moreover, for the class of decomposable indices listed in the main result (Theorem 3) we show that the function that aggregates inequality values in subgroups is necessarily the generalized mean, which in case of the weighted average index reduces to the arithmetical mean (Remark 2).

Among the indices proposed in the literature, we find that the only decomposable index is an index which is called the absolute value index in Abul Naga and Yalcin (2008) and which is also Apouey index with linear function (Apouey 2007). Although Apouey (2007) studies polarization, the proposed indices fulfill the postulated inequality axioms, therefore decomposability can be considered. The Allison and Foster (2004) index is not decomposable and we did not study the decomposability of an inequality measure proposed by Zheng (2010) since it involves socioeconomic inequalities (strictly speaking, there are two relevant dimensions, namely health and socioeconomic status) whereas we deal with pure inequalities in health only. The measures proposed by Blair and Lacy (2000) do not fulfill our definitions of decomposability either. Zheng (2008) points that the measures of Blair and Lacy (2000) and Allison and Foster (2004) should be considered as polarization not inequality indices since they measure only how concentrated the data are around the two ends. Altough the relationship between the concepts of inequality and polarization in case of ordinal data is not the subject of this paper, we notice that Theorems 1 and 2 do not make any use of inequality axioms and hence these decomposability results are independent on this discussion.

Next, we use the Swiss SHRS to calculate the contribution of health inequality of seven Swiss regions to the overall health inequality in Switzerland. We use the absolute value index and the weighted absolute value index for which weights are such that either more weight is attached to below median dispersion or more weight is attached to above median dispersion. Leman is by any measure the most unequal region in Switzerland and Ticino is the least by two measures. However, relatively low variation in SHRS data makes contributions look similar to population sizes, with Middle-Land and Leman contributing the most (respectively, 0.243 and 0.211 by the absolute value index) to overall health inequality.

The paper is organized as follows. In Section 1 we present notation and define axioms. In Section 2 we state characterization theorems that provide explicit functional forms for the class of decomposable indices. In Section 3 we check for the decomposability of the indices already existing in the literature. In Section 4 we present the analysis of health inequality decomposition in Switzerland by population groups which are seven statistical regions. Finally, we conclude.[5]

1 Notation and axioms

Following Abul Naga and Yalcin (2008) we call a vector of n categories $c = (c_1, \ldots, c_i, \ldots, c_n)$ a scale whenever $c_1 < \ldots < c_i < \ldots < c_n$. Let C denote the set of all such ordered increasing scales. It makes sense to work with scales which have at least two categories and consequently in what follows it is assumed that $n \geq 2$. For instance, we have ordered responses to health status and $c = (1, 2, 3, 4, 5)$ means that the first health category is assigned number 1, the second health category is assigned number 2 and the third, the fourth and the fifth categories are assigned, respectively, numbers $3, 4, 5$. Let p_i denote the proportion of individuals in the class c_i. Obviously we require $p_i \in [0, 1]$ and $\sum_{i=1}^{n} p_i = 1$. A

[5]Proofs of the main theorems are available upon request or can be downloaded from http://coin.wne.uw.edu.pl/mkobus/Inequality decomposition by population subgroups for ordinal data.pdf.

frequency distribution and an associated cumulative distribution are, respectively, $\pi := (p_1, \ldots, p_n)$ and $\Pi := (\Pi(c_1), \ldots, \Pi(c_n))$. A cumulative distribution can be also identified with $\Pi = (P_1, \ldots, P_n)$, where $P_i := \sum_{k=1}^{i} p_k = \Pi(c_i)$; further π is an element of λ and Π is an element of Λ, which denote, respectively, the set of all distributions and cumulative distributions defined over n discrete states. We let state m be the median of Π if $P_{m-1} \leq 0.5$ and $P_m \geq 0.5$. Please note that the median does not have to be unique. Let $I : \Lambda \times C \mapsto \mathbb{R}$ be an inequality index for qualitative data.

As we mentioned in the Introduction we will make use of the Allison-Foster (AF forthwith) ordering for evaluating the equality present in the distribution. Formally, $\prec_{AF}$ denotes the partial ordering of the distributions. Let $\Pi, \Omega := (Q_1, \ldots, Q_n)$ be two elements of Λ. We say that $\Pi \prec_{AF} \Omega$ if and only if the following three conditions are met:

(AF1) Π, Ω have identical median states m,

(AF2) $P_i \leq Q_i$ for any $i < m$,

(AF3) $P_i \geq Q_i$ for any $i \geq m$.

The intuition behind Allison-Foster ordering is that Π is more concentrated around the median state than Ω. As Abul Naga and Yalcin (2008) point out AF ordering essentially requires that a transfer in the spirit of the Pigou-Dalton transfer, namely from a person initially above the median to a person below the median and moving both individuals into direction of the median induces a decrease of the inequality index. For example, the cumulative distributions corresponding to distributions π and ω presented in the Introduction are, respectively, $\Pi = (0.2, 0.4, 0.6, 0.8, 1)$ and $\Omega = (0.3, 0.5, 0.6, 0.7, 1)$. The median state of Π is the third category, whereas Ω has two median states: the second and the third category, hence the third category is the common median state. As we see $0.2 < 0.3; 0.4 < 0.5$ and $0.8 > 0.7$, therefore $\Pi \prec_{AF} \Omega$.

Following Abul Naga and Yalcin (2008) we introduce the most equal $\hat{\pi}$ and the most unequal $\check{\pi}$ distribution which are, respectively, the distribution in which all probability mass is concentrated in one category and the distribution in which half of probability mass is concentrated in the lowest category and the other half in the highest category.

The following axioms will be imposed on inequality indices.

CON $I : \lambda \times C \mapsto \mathbb{R}$ is a continuous function.

SCALINV $(I(\pi, c_1) \leq I(\omega, c_1)) \Leftrightarrow (I(\pi, c_2) \leq I(\omega, c_2))$ for any $c_1, c_2 \in C$ and any $\pi, \omega \in \lambda$.

SCALINDEP $(I(\pi, c_1) = I(\pi, c_2))$ for any $c_1, c_2 \in C$ and any $\pi \in \lambda$.

NORM $I(\pi, c) \geq 0$ with $I(\hat{\pi}, c) = 0$ and $I(\check{\pi}, c) = 1$ for any $c \in C$.

EQUAL $(\pi \prec_{AF} \omega) \Rightarrow (I(\pi, c) \leq I(\omega, c))$ for any $c \in C$.

These axioms parallel standard axioms used in inequality measurement. CON states that an index is continuous, whereas SCALINV requires that the ordering of distributions established by an index is invariant to scale changes. In other words, SCALINV ensures that the situation described in the example given in the Introduction cannot happen; that is, whether one distribution exhibits more inequality than the other does not change with the way the numbers are assigned to particular categories. SCALINDEP is even stronger since it makes the index independent of the scale, thus we write $I(\pi)$. Obviously if SCALINDEP holds, then so does SCALINV. NORM requires that the index be normalized i.e. zero is assigned to the most equal distribution and one is assigned to the most unequal distribution. EQUAL states that the index is consistent with the Allison-Foster equality ordering.

Decomposability is defined as follows.

DECOMP There exists a $f : Ran(I) \times Ran(I) \times (0,1) \times C \mapsto \mathbb{R}$ that is continuous and strictly increasing with respect to the first two coordinates such that for any $\pi, \omega \in \lambda, \alpha \in (0,1)$

$$I(\alpha\pi + (1-\alpha)\omega, c) = f(I(\pi, c), I(\omega, c), \alpha, c), \tag{1}$$

where $\alpha\pi + (1-\alpha)\omega$ is a weighted sum of probability distributions (i.e. π assigns mass p_i to category $c_i \in c$ and ω assigns mass q_i, then the probability mass attributed to c_i in $\alpha\pi + (1-\alpha)\omega$ is $\alpha p_i + (1-\alpha)q_i$). If an index fulfills SCALINDEP then (1) becomes

$$I(\alpha\pi + (1-\alpha)\omega) = f(I(\pi), I(\omega), \alpha). \tag{2}$$

DECOMP requires that an index be presented as some function of inequality values in subgroups and subgroup sizes expressed in percentages. In order to better understand how DECOMP works we consider the following example. Let $\pi := (0.25, 0.25, 0.50); \omega := (0.30, 0.40, 0.30)$ and $\alpha = 0.5$. The distribution $0.5\pi + 0.5\omega := (0.275, 0.325, 0.40)$ can be viewed as two population subgroups of equal size $\alpha = 0.5$ that correspond to distributions π and ω. Then, if the inequality index fulfills DECOMP the inequality value associated with the distribution $(0.275, 0.325, 0.40)$ can be decomposed into inequality values in groups π and ω.

Alternatively, one can consider a more general definition of decomposability such that, *for certain fixed* $k \geq 2$ equation (2) becomes

$$I\left(\sum_{i=1}^{k} \alpha_i \pi_i\right) = f(I(\pi_1), \ldots, I(\pi_k), \alpha_1, \ldots, \alpha_k), \tag{3}$$

where $\sum_{i=1}^{k} \alpha_i = 1$.[6]. Therefore, this definition is stronger than DECOMP. Yet in our setting it proves to be equivalent. The same is the case if we require that (3) holds *for every* k. This is explained in Remark 1 below.

2 Characterization theorems

In this section we characterize indices by the axioms introduced in the previous section.

[6]We would like to thank a referee for this alternative definition.

Theorem 1. *I fulfills CON, NORM, SCALINDEP, DECOMP if and only if I is of the form*

$$I(\pi) = G\left(\sum_{i=1}^{n} a_i p_i\right), \tag{4}$$

where $(a_1, a_{2,} \ldots, a_n) \in \mathbb{R}^n$, $G : \mathbb{R} \mapsto [0,1]$ *is a continuous strictly monotonic function. Moreover*

$$G(\hat{x}) = 0 \text{ and } G(\check{x}) = 1,$$

where $\hat{x} = \sum_{i=1}^{n} a_i \hat{p}_i$, $\check{x} = \sum_{i=1}^{n} a_i \check{p}_i$ *and,* $\hat{p}_i$ *and* $\check{p}_i$ *correspond to* $\hat{\pi}$ *and* $\check{\pi}$ *respectively.*

The four properties stated in Theorem 1 are sufficient to reduce the class of considered indices only to continuous and normalized transformations of a weighted average of frequencies. In obtaining functional form (4) DECOMP and SCALINDEP play a crucial role. Although (4) is a natural conjecture, the proof turned out to be quite involved.

Remark 1. *Theorem 1 holds true if we use the alternative definitions of DECOMP e.g. such as condition* (3).

Proof. We start with the "if" part. As we noticed alternative definitions imply DECOMP hence (3) plus the other axioms implies that (4) is the only possible functional form. Now we check the "only if" part. Let $\pi_1, \ldots, \pi_k$ be distributions such that $\pi_l = (p_1^l, \ldots, p_n^l)$. We put $\pi = \sum_{l=1}^{n} \alpha_l \pi_l$ and

$$p_i = \sum_{l=1}^{k} \alpha_l p_i^l.$$

We have

$$I\left(\sum_{l=1}^{k} \alpha_l \pi_l\right) = I(\pi) = G\left(\sum_{i=1}^{n} a_i p_i\right) = G\left(\sum_{i=1}^{n} a_i \sum_{l=1}^{k} \alpha_l p_i^l\right)$$
$$= G\left(\sum_{l=1}^{k} \alpha_l \sum_{i=1}^{n} a_i p_i^l\right) = f\left(I(\pi_1), \ldots, I(\pi_k), \alpha_1, \ldots, \alpha_k\right),$$

where $f(x_1, \ldots, x_k, \alpha_1, \ldots, \alpha_k) = G\left(\sum_{l=1}^{k} \alpha_l G^{-1}(x_l)\right)$. □

To require that the inequality index for ordinal data does not depend on the scale may be considered as too strong a condition and consequently SCALINDEP may appear as too strong an axiom. What is important is that an index does not change arbitrarily with the scale and for this to hold one does not need to impose independence on the scale. As already mentioned the SCALINV axiom ensures the required invariance property of an inequality measure. Therefore, a closely related result to Theorem 1 emerges when we replace SCALINDEP with SCALINV.

Theorem 2. *The index I fulfills CON, SCALINV, NORM and DECOMP if and only if I is of the form*

$$I(\pi, c) = G\left(\sum_{i=1}^{n} a_i p_i, c\right), \tag{5}$$

for some $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ *and* $G : \mathbb{R} \times C \mapsto [0,1]$ *is a continuous strictly monotonic (with respect to the first coordinate) function. Moreover*

$$G(\hat{x}, c) = 0 \text{ and } G(\check{x}, c) = 1, \quad \text{for any } c \in C,$$

where $\hat{x} = \sum_{i=1}^{n} a_i \hat{p}_i, \check{x} = \sum_{i=1}^{n} a_i \check{p}_i$ *and* $\hat{p}_i$ *and* $\check{p}_i$ *correspond to* $\hat{\pi}$ *and* $\check{\pi}$ *respectively.*

Relaxing SCALINDEP by replacing it with SCALINV adds dependence of the index on the scale. Also, for each scale c function $G(\cdot, c)$ is increasing (or, equivalently, we could demand it to be decreasing) with respect to the first coordinate. This is quite intuitive given that SCALINV requires that the ordering imposed by an index is invariant to scale changes.

So far we have not considered EQUAL axiom, which gives us a criterion by which we judge whether one distribution is more equal than the other. Thus the question now to answer is what additional structure is added by this axiom.

Theorem 3. *Let I be an index which can be decomposed according to* (4)*. Then, I fulfills EQUAL if and only if there exist G which fulfills conditions listed in Theorem 1 and* $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$*, such that*

$$I(\pi) = G\left(\sum_{i=1}^{n} a_i p_i\right). \tag{6}$$

Moreover, either G is a strictly increasing function and $a_i \geq a_{i+1}$ when $i < m$ and $a_i \leq a_{i+1}$ when $i \geq m$ or G is a strictly decreasing function and $a_i \leq a_{i+1}$ when $i < m$ and $a_i \geq a_{i+1}$ when $i \geq m$.

EQUAL axiom states that the distribution that is more concentrated around its median state is more equal, or equivalently, the less concentration around the median state, the more inequality. Hence, index (6) assigns more weight to categories which are further from the median.

Knowing that an index has the form (6), we may write decomposition (1) explicitly.

Remark 2. *Let I be given by* (6). *We denote $\tilde{I}(\pi) := \sum_{i=1}^{n} a_i p_i$ which is an index for which* (1) *writes as*

$$\tilde{I}(\alpha\pi_1 + (1-\alpha)\pi_2) = \alpha\tilde{I}(\pi_1) + (1-\alpha)\tilde{I}(\pi_2), \tag{7}$$

i.e. f is the arithmetical mean given by $f(i_1, i_2, \alpha) = \alpha i_1 + (1-\alpha) i_2$.

By definition $I = G \circ \tilde{I}$, therefore we have

$$\begin{aligned} I(\alpha\pi_1 + (1-\alpha)\pi_2) &= G(\alpha\tilde{I}(\pi_1) + (1-\alpha)\tilde{I}(\pi_2)) \\ &= G\left(\alpha(G^{-1} \circ I)(\pi_1) + (1-\alpha)(G^{-1} \circ I)(\pi_2)\right). \end{aligned}$$

i.e. f is the generalized mean

$$f(i_1, i_2, \alpha) = G\left(\alpha G^{-1}(i_1) + (1-\alpha)G^{-1}(i_2)\right)$$

which in case of $G(\pi) = \pi$ reduces to the arithmetical mean.

3 Decomposability of specific indices

Based on Theorems 1-3 we now check which of the indices proposed in the literature are decomposable.

- *Absolute value index*

This index was proposed by Abul Naga and Yalcin (2008):[7]

$$I_{1,1} = \frac{\sum_{i<m} P_i - \sum_{i\geq m} P_i + (n+1-m)}{(n-1)/2} \tag{8}$$

We will now present $I_{1,1}$ in the form given by Theorem 3. We have

$$\begin{aligned}\sum_{i<m} P_i - \sum_{i\geq m} P_i &= \sum_{i=1}^{m-1}\left(\sum_{j=1}^{i} p_j\right) - \sum_{i=m}^{n}\left(\sum_{j=1}^{i} p_j\right) \\ &= \sum_{i=1}^{m-1}(m-i)p_i - (n-m+1)\sum_{i=1}^{m-1} p_i - \sum_{i=m}^{n}(n-i+1)p_i \\ &= \sum_{i=1}^{m-1}(2m-n-i-1)p_i + \sum_{i=m}^{n}(-n+i-1)p_i.\end{aligned}$$

Clearly, for $i < m$ weights are decreasing with i while for $i \geq m$ they are increasing hence $I_{1,1}$ is of the form (6). Absolute value index is a member of the family of indices $I_{\alpha,\beta}$ (Abul Naga and Yalcin, 2008) for which $\alpha = \beta = 1$; however in general the indices that belong to $I_{\alpha,\beta}$ are not decomposable.

One can think of the following generalization which gives us the weighted absolute value index.

$$I_{a,b} = \frac{a\sum_{i<m} P_i - b\sum_{i\geq m} P_i + b(n+1-m)}{\left(a(m-1) + b(n-m)\right)/2}; \quad a, b \geq 0. \tag{9}$$

If $a = 1$ and $b = 1$, then obviously we get $I_{1,1}$ and it is evident that adding these weights does not change decomposability property. There is a nice interpretation related to the weighted absolute value index. When $a > b$ the index is more sensitive to inequality below the median, whereas the opposite is true if $a < b$ and more weight is attached to inequality above the median.

- *Allison-Foster index*

The following index was proposed by Allison and Foster (2004):

$$I_{AF} = \frac{\sum_{i=m}^{n} c_i p_i}{\sum_{i=m}^{n} p_i} - \frac{\sum_{i=1}^{m-1} c_i p_i}{\sum_{i=1}^{m-1} p_i}$$

Because of changing weights it is neither of the form (6) nor (5).

[7]This is an index denoted as $I_{1,1}$ in Abul Naga and Yalcin (2008), however here the subscript $\{1,1\}$ means that weights are equal to one (see the weighted index defined later in the text).

- *Apouey index*

  This is the index of Apouey (2007):

$$I_A := C_1 \sum_{i=1}^{n} g\left(\left|P_i - \frac{1}{2}\right|\right) + C_2,$$

where g is a continuous decreasing function and $C_1 > 0$. Obviously, I_A fulfills CON, by Proposition 1 in Apouey (2007) it is consistent with EQUAL. Applying NORM when $g(x) = x$ we know that $I_A = I_{1,1}$. For $g(x) = x$ using the properties of the modulus function we have

$$I_A = C_1 \left(\sum_{i \geq m} P_i - \sum_{i<m} P_i + m - \frac{n}{2} - 1 \right) + C_2$$

The values of I_A calculated for, respectively, the best and the worst distributions give us the following system of equations

$$C_1 \frac{n}{2} + C_2 = 0 \quad \text{and} \quad C_1 \frac{1}{2} + C_2 = 1,$$

thus we get

$$C_1 = \frac{2}{1-n} = -\frac{2}{n-1}, \quad C_2 = \frac{n}{n-1}$$

Substituting C_1, C_2 into the expression for I_A after simple calculations we get $I_{1,1}$. Let us also notice that any other linear function $g(x) = ax+b$, $a > 0$ will also induce $I_{1,1}$ (obviously with different C_1, C_2).

The above is the only case when the index is decomposable. We will now present a sketch of the proof of this fact. Let us assume that I_A can be decomposed, i.e.

$$\sum_{i=1}^{n} g\left(\left|P_i - \frac{1}{2}\right|\right) = G\left(\sum_{i=1}^{n} a_i p_i\right), \tag{10}$$

for some G and a_i's (we skip C_1 and C_2 for notational convenience). This resembles a little bit the Pexider equation. Following this path, we notice that making a few simple calculations we can find $\tilde{a}_i$'s such that $\sum_{i=1}^{n} a_i p_i = \sum_{i=1}^{n} \tilde{a}_i P_i$. We also denote $h_i(x) = g(|x/a_i - 1/2|)$. Let us fix $P_2, P_3, \ldots, P_n$. We could write (10) as

$$h_1(x_1) + h_2(x_2) = \tilde{G}(x_1 + x_2),$$

where one should think about x_i as $\tilde{a}_i P_i$ and $\tilde{G}$ is defined in an obvious way. By (Aczel, 1966, Theorem 1 p. 142) and assumption of continuity of g (and consequently of h_i's) we obtain that h_i's have to be linear. This proves the linearity of g, potentially only on some subset of $[0, 1/2]$. Varying P_2 we could extend it to the whole $[0, 1/2]$ (we note that the behavior outside of $[0, 1/2]$ does not influence I_A). The details are left to the reader.

- *Blair and Lacy index*

The indices of Blair and Lacy (2000) are the following:

$$I_{BL} := 1 - \frac{\sum_{i=1}^{n-1} (P_i - 0.5)^2}{(n-1)/4}$$

and

$$\hat{I}_{BL} := 1 - \left(\frac{\sum_{i=1}^{n-1} (P_i - 0.5)^2}{(n-1)/4} \right)^{\frac{1}{2}}$$

These indices fulfill CON, NORM and SCALINDEP but they are not decomposable. To see this let us assume that I in (4) is differentiable and let us calculate its gradient

$$\nabla I(\pi) = \left(\frac{\partial I(\pi)}{\partial p_1}, \frac{\partial I(\pi)}{\partial p_2}, \dots, \frac{\partial I(\pi)}{\partial p_n} \right) = G' \left(\sum_{i=1}^{n} a_i p_i \right) (a_1, a_2, \dots, a_n).$$

One easily notices that gradients for any two points are collinear. We will check that the gradient of I_{BL} does not have this property. Indeed,

$$\frac{\partial I_{BL}}{\partial p_j} = -\frac{8}{n-1} \sum_{i=j}^{n-1} (P_i - 0.5).$$

Now it is obvious that gradient of I_{BL} does not have the above property. For example, one can take two cdf's $(0.1, 0.2, 0.3, 1, \dots, 1)$ and $(0.1, 0.1, 0.4, 1, 1, \dots, 1)$.

Calculation of the gradient of $\hat{I}_{BL}$ is substantially harder so instead we notice that $\hat{I}_{BL} = H \circ I_{BL}$, where $H(x) = 1 - (1-x)^{1/2}$. Moreover DECOMP is invariant with respect to the monotonic transformations (which H is). Since I_{BL} is not decomposable using Theorem 1 we notice that the only axiom that is not fulfilled is DECOMP. Hence $\hat{I}_{BL}$ cannot fulfill DECOMP either.

Table 1: Distribution of SHRS in the seven statistical regions of Switzerland

Area	Population	SHRS distribution				
	%	Very bad	Bad	So so	Good	Very good
Leman	18	0.01	0.05	0.16	0.72	1
North-West	14	0.01	0.05	0.18	0.81	1
Central	9	0.00	0.02	0.13	0.76	1
Middle-Land	23	0.01	0.04	0.17	0.77	1
East	15	0.00	0.03	0.14	0.78	1
Ticino	4	0.01	0.06	0.17	0.87	1
Zurich	17	0.00	0.03	0.13	0.78	1

Source: Abul Naga and Yalcin (2008) and Eurostat database.

4 Empirical application

Based on the data concerning 2002 wave of Swiss health survey provided in Abul Naga and Yalcin (2008) we evaluate the impact of health inequality in seven statistical regions of Switzerland on the overall inequality in Switzerland. Table 1 presents distribution of SHRS and population contributions for the year 2002 for seven Swiss regions.

We need to calculate the impact of inequality in seven subgroups on the total inequality. Clearly, for k subgroups, by induction, (7) reads as follows

$$I(\alpha_1\pi_1 + \alpha_2\pi_2 + \ldots + \alpha_k\pi_k) = \alpha_1 I(\pi_1) + \alpha_2 I(\pi_2) + \ldots + \alpha_k I(\pi_k). \tag{11}$$

and the same holds if $\tilde{I}$ depends on P_i instead of π_i. Here $k = 7$ and α's are regions' population sizes (percentages). The median of the SHRS distribution in every region is category fourth labeled "good". As inequality indices we use the absolute value index and its weighted version. For $m = 4$ and $n = 5$ these indices are the following

$$I_{1,1} = \frac{\sum_{i<m} P_i - \sum_{i \geq m} P_i + 2}{2}$$

Table 2: Inequality decomposition by population subgroups in Swiss regions

Area	$I_{2,1}$	Contribution	$I_{1,1}$	Contribution	$I_{1,2}$	Contribution
Leman	0.205	0.207	0.250	0.211	0.312	0.215
North-West	0.191	0.149	0.215	0.141	0.248	0.133
Central	0.154	0.078	0.195	0.082	0.252	0.087
Middle-Land	0.191	0.246	0.225	0.243	0.272	0.239
East	0.160	0.134	0.195	0.137	0.244	0.140
Ticino	0.174	0.039	0.185	0.035	0.200	0.030
Zurich	0.154	0.147	0.190	0.151	0.240	0.156

and

$$I_{a,b} = \frac{a\sum_{i<m} P_i - b\sum_{i\geq m} P_i + 2b}{(3a+b)/2}.$$

In what follows we admit the following weights: $a = 2, b = 1$; $a = 1, b = 1$ and $a = 1, b = 2$. Obviously, $a = 1, b = 1$ gives us the absolute value index. Total inequality as measured by these three different indices is the following: $I_{2,1} = 0.178943, I_{2,1} = 0.21335, I_{2,1} = 0.26152$.

Let us now study regions' contributions to overall inequality (Table 2). The contribution of region k is calculated according to $\frac{\alpha_k I(\pi_k)}{I(\pi)}$, where α_k is region k's population size (or more precisely, percentage of the overall population attributed to region k) and $I(\pi_k), I(\pi)$ are inequality values in, respectively, region k's SHRS distribution and overall distribution. Three inequality rankings (beginning from the highest inequality) are the following:

$$\text{Leman} >_{I_{2,1}} \text{North-West} =_{I_{2,1}} \text{Middle-Land} >_{I_{2,1}} \text{Ticino} >_{I_{2,1}} \text{East} >_{I_{2,1}} \text{Central} =_{I_{2,1}} \text{Zurich}.$$

$$\text{Leman} >_{I_{1,1}} \text{Middle-Land} >_{I_{1,1}} \text{North-West} >_{I_{1,1}} \text{Central} =_{I_{1,1}} \text{East} >_{I_{1,1}} \text{Zurich} >_{I_{1,1}} \text{Ticino}.$$

$$\text{Leman} >_{I_{1,2}} \text{Middle-Land} >_{I_{1,2}} \text{Central} >_{I_{1,2}} \text{North-West} >_{I_{1,2}}$$
$$>_{I_{1,2}} \text{East} >_{I_{1,2}} \text{Zurich} >_{I_{1,2}} \text{Ticino}.$$

By any measure Leman is the most unequal region. As to the least unequal region the three indices are not fully consistent, however. It appears that when weight is shifted away from below median inequality, Ticino emerges as having the lowest health inequality. On the other hand, Ticino is the fourth most unequal region according to $I_{2,1}$, which suggests that most inequality in Ticino (in comparison to other regions) occurs at the bottom of the distribution. As we increase the sensitivity of an index to above median categories, the dispersion of inequality values increases; that is, the dispersion for $I_{2,1}$ is 0.051 and for $I_{1,2}$ the dispersion equals 0.112. This implies that health distributions differ at most in higher categories, which is consistent with the findings of Abul Naga and Yalcin (2008).

The contribution rankings are the following:

$$\text{Middle-Land} >_{I_{2,1}} \text{Leman} =_{I_{2,1}} \text{North-West} >_{I_{2,1}} \text{Zurich} >_{I_{2,1}}$$
$$>_{I_{2,1}} \text{East} >_{I_{2,1}} \text{Central} =_{I_{2,1}} \text{Ticino}.$$

$$\text{Middle-Land} >_{I_{1,1}} \text{Leman} >_{I_{1,1}} \text{Zurich} >_{I_{1,1}} \text{North-West} =_{I_{1,1}}$$
$$=_{I_{1,1}} \text{East} >_{I_{1,1}} \text{Central} >_{I_{1,1}} \text{Ticino}.$$

$$\text{Middle-Land} >_{I_{1,2}} \text{Leman} >_{I_{1,2}} \text{Zurich} >_{I_{1,2}} \text{East} >_{I_{1,2}}$$
$$>_{I_{1,2}} \text{North-West} >_{I_{1,2}} \text{Central} >_{I_{1,2}} \text{Ticino}.$$

The highest contribution to total health inequality in Switzerland is attributed to the Middle Land and the second highest is attributed to Leman, irrespectively of the inequality measure. Thus judging inequality contribution only on the basis of inequality value would be misleading. Ticino is the region that contributes the least and this is caused by its smallest population size as well as its low inequality score. The last two rankings are identical to the ordering of population sizes, which indicates that apparently variation in health response data is suppressed by relatively large differences in populations sizes. In the first ranking, when more weight is put at the lower end of the distribution (relatively small) difference in population sizes between North West, East and Zurich is not enough to compensate for higher percentage of individuals with bad health status in North West and consequently the contribution of North West is greater than the contributions of both East and Zurich.

Conclusions

This paper addresses the problem of inequality decomposition by population subgroups for qualitative data such as health surveys. Conventional inequality measures are not well-suited to handle qualitative data. As the main result we derived an explicit functional form for inequality indices that satisfy decomposability along with other standard inequality axioms. We applied our methodology to the study of health inequality decomposition in Switzerland by groups that consist of seven statistical regions. Inequality decomposition into groups defined by race, sex, gender can be carried out in the same manner. Our empirical example focused on health data, yet other outcome variables can also be considered. Although decomposability itself is a desired property of inequality measures, it turns out that it implies a severe restriction on the form of inequality indices. Namely, decomposable indices are necessarily non-decreasing transformations of an index which is a weighted average of frequencies in considered categories.

References

1. Allison R A, Foster J E. Measuring health inequality using qualitative data, Journal of Health Economics 2004;23(3); 505-524.

2. Abul Naga R H, Yalcin T. Inequality measurement for ordered response health data, Journal of Health Economics 2008;27(6); 1614-1625.

3. Aczel J. Lectures on functional equations and their applications. Academic Press: New York; 1966. 4. Apouey B. Measuring health polarization with self-assessed health data, Health Economics 2007;16; 875-894.

5. Blair J, Lacy M G. Statistics of ordinal variation, Sociological Methods and Research 2000; 28(251);251-280.

6. Shorrocks A F. Inequality decomposition by population subgroups, Econometrica 1984;52(6); 1369-85.

7. Zheng B. A new approach to measure socioeconomic inequality in health, Journal of Economic Inequality 2011:555-577.

8. Zheng B. Measuring inequality with ordinal data: a note, Research on Economic Inequality 2008; 16:177-188.

Faculty of Economic Sciences
University of Warsaw
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl