EASYPol
On-line resource materials for policy making

ANALYTICAL TOOLS

Module 052

Policy Impacts on Inequality

Decomposition of Income Inequality by Subgroups

EASYPol
On-line resource materials for policy making

Policy Impacts on Inequality

Decomposition of Income Inequality by Subgroups

by

Lorenzo Giovanni Bellù, Agricultural Policy Support Service, Policy Assistance Division, FAO, Rome, Italy

Paolo Liberati, University of Urbino, "Carlo Bo", Institute of Economics, Urbino, Italy

for the

Food and Agriculture Organization of the United Nations, FAO

Table of contents

1. SUMMARY

This tool illustrates how to decompose inequality measures by subgroups of populations. In particular, it defines the concepts of within and between inequality and analyses how different inequality indexes perform with respect to this decomposition. In particular, the performance of the analysis of variance, the Gini Index and the Theil Index will be discussed. A step-by-step procedure and numerical examples give operational content to the tool.

2. INTRODUCTION

Objective

The objective of the tool is to provide the analytical and the practical framework to understand where inequality comes from. To this purpose, decomposition of the most suitable indexes will be provided.

Decomposing inequality indexes means exploring the **structure of inequality**, i.e. the disaggregation of total inequality in relevant factors.

For empirical applications, the knowledge of overall inequality may be insufficient to properly target public policies. Actual policies may have a very differentiated impact on subgroups of population (e.g. rural and urban households). It is therefore essential to split overall inequality among different groups of population.

Target audience

This module targets current or future policy analysts who want to increase their capacities in analysing impacts of development policies on inequality by means of income distribution analysis. On these grounds, economists and practitioners working in public administrations, in NGOs, professional organisations or consulting firms will find this helpful reference material.

Required background

Users should be familiar with basic notions of mathematics and statistics.

Links to relevant EASYPol modules, further readings and references are included both in the footnotes and in section 7.2 of this module[1].

[1] EASYPol hyperlinks are shown in blue, as follows:
a) training paths are shown in **underlined bold font;**
b) other EASYPol modules or complementary EASYPol materials are in ***bold underlined italics***;
c) links to the glossary are in **bold**; and
d) external links are in *italics*.

3. CONCEPTUAL BACKGROUND

In all EASYPol modules inequality has been investigated as the «**overall inequality**» among a given set of individuals with given income levels. However, inequality may stem from different groups or sectors of population with different intensities (e.g. workers and pensioners; rural and urban households; households with and without children). A very important feature of inequality measures is therefore **DECOMPOSABILITY**, i.e. the possibility of calculating the contribution of each group to total inequality. In this paragraph we will focus on decomposition by groups of population.

In its most general form, decomposability of inequality measures requires a consistent relation between overall inequality and its parts. More specifically, when dealing with decomposability, we must be able to distinguish between **WITHIN INEQUALITY** (W) and **BETWEEN INEQUALITY** (B). The «within inequality» element captures the inequality due to the variability of income within each group, while the «between inequality» captures the inequality due to the variability of income across different groups.

For example, if the population is divided in urban and rural individuals, the **W** element identifies the contribution to inequality of the variability of urban and rural incomes taken separately. The **B** element, instead, captures the degree of inequality due to income differences between groups.

The most general decomposition of any inequality index *I* generates a **within** element, a **between** element and a **residual term** :

$$I = \underbrace{I_{\mathrm{WIT}}}_{\mathrm{WITHIN}} + \underbrace{I_{\mathrm{BET}}}_{\mathrm{BETWEEN}} + \underbrace{K}_{\mathrm{RESIDUAL}}$$

A natural starting point for the analysis of inequality decomposition is to focus on the **analysis of variance**. The analysis of variance is based on the decomposition of within and between elements. We have also learnt that the variance itself may be used as an inequality index[2].

3.1 The analysis of variance

Let us assume an income distribution with two groups, individuals in urban areas (**U**) and individuals in rural areas (**R**). Denote their incomes as y_i^U and y_i^R, where the subscript refers to a generic individual, while the superscript identifies the area the individual belongs to. In a general form, the variance of total income might be decomposed as follows:

[1] $$V(y) = \underbrace{\left[w_U V(\mathbf{y}_U) + w_R V(\mathbf{y}_R)\right]}_{\mathrm{WITHIN}} + \underbrace{V\left[\bar{y}_U, \bar{y}_R\right]}_{\mathrm{BETWEEN}}$$

[2] See EASYPol Module 080: ***Policy Impacts on Inequality: Simple Inequality Measures***.

The first term is the sum between two elements:

- the **variance of urban incomes** $V(y_U)$ multiplied by the share of urban residents on total population (w_U);

- the **variance of rural incomes** $V(y_R)$ multiplied by the share of rural residents on total population (w_R);

The first term is therefore the **WITHIN** element of the variance and it can be interpreted as the weighted average of the variance of the income of each group, with weights given by the population shares.

Whereas, to understand the second term of [1], imagine calculating mean urban income and mean rural income. Then, replace each actual income with the **mean income of the group the individual belongs to**. The last term is the variance of this fictitious income distribution. As within groups, incomes are all equal and they only differ, possibly, between groups, the variance will pick up the dispersion of income attributable to the difference between groups. This is why this part is called the BETWEEN element of the variance. Note that expression [1] does not contain a residual element. This means that the variance is perfectly decomposable.

3.2 The Gini Index

In order to make decomposability attractive, it is necessary to explore whether and how inequality indexes satisfy this property. We will focus on two indexes: the Gini Index, as this is one of the most popular inequality index, and the family of Theil Indexes.

The Gini Index is the most popular index of inequality[3]. Let us now examine how it performs with respect to the decomposability issue.[4]

It is first worth noting that the Gini Index is not perfectly decomposable, as it has a **non-zero residual *K*** besides the within and between inequality.

Assuming again two groups, rural I and urban (U) individuals, the **within** element of the Gini Index (G_{WIT}) is given by the following formula:

[2] $$G_{\text{WIT}} = \left(\frac{n_U}{n}\frac{Y_U}{Y}\right)G_U + \left(\frac{n_R}{n}\frac{Y_R}{Y}\right)G_R$$

where G_U and G_R are the Gini indexes measured on urban and rural incomes, respectively, while the round brackets are the weights given to each group. These weights are in turn given by the product between the population share of each group (the first term in the round brackets) and the income share of each group (the second term in the round brackets). The logic is quite the same as in the case of variance *V*. To

[3]See EASYPol Module 040: ***Inequality Analysis: The Gini Index***.

[4] Essential reference on this topic is Lambert and Aronson (1993).

decompose the within group element, we must always calculate subgroup indexes and multiply them for the corresponding weights.

The **between** element of the Gini Index (G_{BET}) is calculated as in the case of variance. Indeed, G_{BET} must be calculated on an income distribution where actual incomes are replaced by subgroup mean incomes. Using the covariance formula of the Gini Index, G_{BET} is given by:

[3]
$$G_{\mathrm{BET}} = \frac{2}{\bar{y}} Cov\left(\bar{y}, F(\bar{y})\right)$$

where $\bar{y}$ is the distribution of income obtained by replacing actual incomes with subgroup means.

What the decomposition of the Gini Index adds to the general decomposition [1] is the residual term K. The meaning of this term is not very intuitive, yet its understanding is important to realize when the Gini Index can be used to meaningfully decompose income inequality.

In general, the Gini Index is perfectly decomposable (i.e. K=0) when ranking by subgroup incomes from the poorest to the richest **do not overlap**, i.e. the relative position of each individual is the same as in the total income distribution. The residual K is positive, instead, when ranking by subgroup incomes overlaps, i.e. when the relative position of a given individual in the subgroup income distribution differ from its position in the total income distribution. In what follows, we will illustrate how the decomposition of the Gini Index may be easily interpreted in terms of Lorenz Curves.

3.3 The Theil Index

In the context of additive decomposability, the generalised entropy class of inequality indexes is a good alternative to the Gini Index. Unlike the Gini Index, the members of this class are perfectly decomposable without a residual term. Their economic interpretation is therefore straightforward. Yet, seen from another perspective, they conceal something that the Gini Index can reveal, i.e. the amount of inequality due to the re-ranking effect.

Let us start from the most common member of the GE class, the Theil Index[5]. Assuming m groups, its decomposition assumes the following form:

[4]
$$T = \underbrace{\sum_{k=1}^{m}\left(\frac{n_k}{n}\frac{\bar{y}_k}{\bar{y}}\right)T_k}_{\text{WITHIN}} + \underbrace{\sum_{k=1}^{m}\frac{n_k}{n}\left(\frac{\bar{y}_k}{\bar{y}}\right)\ln\left(\frac{\bar{y}^k}{\bar{y}}\right)}_{\text{BETWEEN}}$$

The first term in [4] is the weighted average of the Theil inequality indexes of each group (T_k), with weights represented by the total income share (the product of

[5] See EASYPol Module 051: ***Policy Impacts on Inequality: The Theil Index and the Other Entropy Class Inequality Indexes.***

population shares and relative mean incomes). This is the **WITHIN** part of the decomposition.

The second term is the Theil Index calculated using subgroup means $\bar{y}_k$ instead of actual incomes. This follows the logic of replacing actual income distributions in each group with the average income level of the same group. This gives the **BETWEEN** part of the decomposition.

Now, the Theil Index is a member of the **GENERALISED ENTROPY CLASS** (GE) of inequality indexes. All members of this family are perfectly decomposable in within and between elements. Recalling the general formula of the GE class, the decomposition can be expressed as follows:

[5]
$$GE_\alpha = \underbrace{\sum_{k=1}^{m}\left(\frac{\bar{y}_k}{\bar{y}}\right)^{\alpha}\left(\frac{n_k}{n}\right)^{1-\alpha} GE(\alpha)_i}_{\text{WITHIN}} + \underbrace{\frac{1}{\left(\alpha^2-\alpha\right)}\left[\sum_{k=1}^{m}\frac{n_k}{n}\left(\frac{\bar{y}_k}{\bar{y}}\right)^{\alpha} - 1\right]}_{\text{BETWEEN}}$$

where now relative means and population shares are raised at α and (1-α) respectively, while $GE(\alpha)_i$ is the GE Index of the *i*-th subgroup. The other terms have the usual meaning.

The first term is again the **WITHIN** part. It is the weighted average of GE Indexes for each group. The second term is the **BETWEEN** element. As usual, it is calculated as a GE Index where actual incomes are replaced by subgroup means, in order to pick up variability only among groups and not within them.

Choosing the desired value of α gives decompositions for the members of the GE class. An interesting result is obtained with α=0, which gives the **mean logarithmic deviation**[6]. In this case the decomposition is the following:

[6]
$$GE_0 = \underbrace{\frac{n_k}{n} GE_0^k}_{\text{WITHIN}} + \underbrace{\frac{n_k}{n}\ln\left(\frac{\bar{y}}{\bar{y}_k}\right)}_{\text{BETWEEN}}$$

In what follows, the focus will be on the Theil Index, as it is the member of GE class most often used to decompose inequality.

[6] See EASYPol Module 080: ***Policy Impacts on Inequality: Simple Inequality Measures.***

4. STEP-BY-STEP PROCEDURE TO DECOMPOSE INEQUALITY

4.1 A step-by-step procedure for the analysis of variance

Figure 1 reports the required steps to decompose inequality by the analysis of variance.

Step 1 asks us to identify the groups from the original income distribution, while Step 2 asks us to sort incomes within each group, in order to have subgroup income distributions ranked by income levels.

Figure 1: A step-by-step procedure for the analysis of variance

STEP	Operational content
1	From the original income distribution, we must identify incomes belonging to the different groups (e.g. rural and urban incomes)
2	Sort incomes within each group
3	Calculate the share of each group in total population
4	Calculate the variance of income for each group separately taken
5	Multiply each variance for the share of the corresponding group in total population
6	Sum all the terms in Step 5. This is the WITHIN element
7	Calculate mean incomes for each group
8	Replace actual incomes of the group with the corresponding means
9	Calculate the variance of this fictitious income distribution. This is the BETWEEN element

Step 3 asks us to calculate the share of each group in total population, i.e. how many people of a given group there are in the total population. This is a share of each group in total population and **not** in total income.

In Step 4, we must calculate the variance of income for each separate group. In Step 5, these variances must be multiplied by the share of each group in the total population as calculated in Step 3. The sum of all these terms give the **within** element (Step 6).

To calculate the **between** element, we must first calculate mean incomes for each group (Step 7). These mean incomes must be used to replace actual incomes in order to create a fictitious income distribution where all members of each group have the same mean income (Step 8). The variance of this simulated income distribution is the between element (Step 9).

4.2 A step-by-step procedure to decompose the Gini Index

To decompose inequality as measured by the Gini Index, we can follow the steps reported in Figure 2.

Step 1 and Step 2 are the same as in the case of variance. Extremely important is that the cumulative distribution function assigned to this distribution is also calculated within each subgroup i.e. instead of having a unique $F(y)$ we have as many $F_i(y)$ as there are groups, as if they were distinct income distributions.

Step 3 now asks us to calculate the population share $\left(\frac{n_i}{n}\right)^2$ and the income share of each group $\left(\frac{y_i}{y}\right)$.

In step 4, we must multiply the income share of a group by the respective squared population share to obtain the weight of each group.

In Step 5, we must calculate the Gini Index of each subgroup income distribution. After that, each Gini Index must be multiplied by the corresponding weight (as calculated in step 4) to obtain the contribution given by each group to G_{WIT} (Step 6). The sum of all groups' contributions gives the **WITHIN** element (Step 7).

To calculate the between element, instead, there are no conceptual differences with the analysis of variance. Step 8 asks us to calculate subgroup mean incomes, while Step 9 asks us to replace actual incomes of each subgroup with the corresponding means. Finally, in Step 10, we must calculate the Gini Index on this simulated income distribution keeping the same cumulative distribution function $F(y)$ as the original income distribution and **not** the subgroup cumulative distribution function. This gives the **BETWEEN** element.

The step-by-step procedure might end here if there would be no residual K, i.e. when ranking by subgroup incomes do not overlap. If these rankings overlap, the decomposition is not perfect and the residual K is positive. In order to calculate the residual, a further step is required: calculate the Gini Index of the original income distribution and subtract the sum of G_{BET} and G_{WIT} so far obtained (Step 11).

Figure 2: A step-by-step procedure to decompose the Gini Index

STEP	Operational content
1	From the original income distribution, we must identify incomes belonging to the different groups (e.g. rural and urban incomes)
2	Sort incomes within each group
3	Calculate the share of each group in total population and the share of each group in total income
4	Multiply the income share by the population share. This gives the full weight of each group.
5	Calculate the Gini index of incomes of each group taken separately
6	Multiply each Gini index for the corresponding weight. It gives the contribution of each group to the WITHIN element
7	Sum all contributions as in Step 6. This gives the WITHIN element
8	Calculate mean incomes for each group
9	Replace actual incomes of the group with the corresponding means
10	Calculate the GINI of this fictitious income distribution. This is the BETWEEN element
11	Calculate the GINI of the original income distribution and subtract both the WITHIN and the BETWEEN element. This gives K

4.3 A step-by-step procedure to decompose the Theil Index

Figure 3 reports the steps needed to get a decomposition of the Theil Index. There are no meaningful differences with the decomposition of other indexes. As usual, we must first identify groups of population and to sort incomes within each group (Step 1 and Step 2).

Then, the share of each group in total income must be calculated (Step 3). Step 4 asks us to calculate the Theil Index of each group separately taken. This index must then be multiplied by the income share (Step 5) in order to identify the contribution of each group to the Within inequality. The sum of all these contributions gives the **Within** element (Step 6).

To calculate the **Between** element, just proceed as usual from Step 7 to 9, by building a simulated income distribution by replacing actual incomes with mean subgroup incomes. It is worth checking the exactness of the decomposition (Step 10).

Figure 3: A step-by-step procedure to decompose the Theil Index

STEP	Operational content
1	From the original income distribution, you must identify incomes belonging to the different groups (e.g. rural and urban incomes)
2	Sort incomes within each group
3	Calculate the share of each group in total income
4	Calculate the Theil Index of incomes of each group taken separately
5	Multiply each Theil Index for the corresponding income share. It gives the contribution of each group to the WITHIN element
6	Sum all contributions as in Step 5. This gives the WITHIN element
7	Calculate mean incomes for each group
8	Replace actual incomes of the group with the corresponding means
9	Calculate the Theil Index of this fictitious income distribution. This is the BETWEEN element
10	Check for the decomposition

5. A NUMERICAL EXAMPLE OF HOW TO DECOMPOSE INEQUALITY INDEXES

5.1 An example of how to perform the analysis of variance

Table 1 reports an example of how to perform an analysis of variance according to the step-by-step procedure discussed in the previous paragraphs.

Steps 1 and 2, in Table 1, distinguish between rural and urban individuals in the usual income distribution. For simplicity, rural individuals have the lowest three incomes. The share of rural population is 43 per cent, while the share of urban residents is 57 per cent. The corresponding variances are 1,815,556 for rural incomes and 2,695,000 for urban incomes (Steps 3 and 4).

Table 1: Analysis of variance

STEP 1 and 2

From the original income distribution, identify incomes belonging to different groups. Sort incomes within each group

Individuals	Incomes	Type (*)
1	2,000	R
2	4,300	R
3	5,200	R
4	8,500	U
5	8,800	U
6	11,000	U
7	12,500	U

(*) R=Rural; U=Urban

STEP 3 and 4

Calculate the share of each group in total population. Calculate the variance of income for each group separately taken

Share of rural	0.43
Share of urban	0.57
Variance of rural	1,815,556
Variance of urban	2,695,000

STEP 5 and 6

Multiply each variance for the share of each group in total population. Sum all these terms to get the WITHIN element

Rural	778,095
Urban	1,540,000
WITHIN	**2,318,095**

STEP 7 and 8

Calculate mean incomes for each group and replace actual incomes with the corresponding means

Individuals	Mean Incomes	Type (*)
1	3,833	R
2	3,833	R
3	3,833	R
4	10,200	U
5	10,200	U
6	10,200	U
7	10,200	U

Mean income rural	3,833
Mean income urban	10,200

STEP 9

Calculate the variance of the income distribution of Step 8. This is the BETWEEN element

BETWEEN	**9,926,803**

CHECK:

Original variance	12,244,898
WITHIN INEQUALITY	2,318,095
BETWEEN INEQUALITY	9,926,803
TOTAL INEQUALITY	12,244,898

Multiplying the variance of incomes of each group for the corresponding share and adding these terms we get the **WITHIN** element (2,318,095) (Steps 5 and 6).

To calculate the BETWEEN element, we must first replace actual incomes of each group by the corresponding means (3,833 and 10,200 for rural and urban incomes, respectively). This is done in Step 7 and 8, where a new income distribution is built. Note that all members of the same group have the same income, while income differs between groups. It means that the variability of income within group is now zero. Therefore, the variance of this simulated income distribution records only the variability of income due to the fact that income between groups varies. The calculated **BETWEEN** element is 9,926,803 (Step 9).

Now, it is worth checking for the exactness of this decomposition. The variance of the original income distribution is 12,244,898. The sum of the two elements gives the same result. Which kind of information did we gain in decomposing inequality? That total inequality is due partly to the variability of income within groups (18.9 per cent, equal to the ratio between 2,318,095 and 12,244,898) and partly to the variability of income between groups (81.1 per cent, equal to the ratio between 9,926,803 and 12,244,898). Therefore, the greatest part of total inequality is explained by how incomes vary between groups, while the remaining part is due to how income varies within each group.

5.2 An example of how to decompose the Gini Index

A numerical example of how to decompose the Gini Index is reported in Table 2.

Table 2: Decomposing the Gini Index without residual

STEP 1

From the original income distribution, identifiy incomes belonging to different groups

Cumulative distribution function	Incomes	Type (*)
0.143	2,000	R
0.286	4,300	R
0.429	5,200	R
0.571	8,500	U
0.714	8,800	U
0.857	11,000	U
1.000	12,500	U
Total income	52,300	
Mean income	7,471	

STEP 2

Sort incomes in each subgroup

Subgroup cumulative distribution function	Incomes	Type (*)
0.3333	2,000	R
0.6667	4,300	R
1.0000	5,200	R
0.2500	8,500	U
0.5000	8,800	U
0.7500	11,000	U
1.0000	12,500	U

STEP 3, 4, 5, 6 and 7

Calculate population share and income share. Calculate subgroup Gini Indexes and multiply them by shares. This gives contributions to G(WIT). Sum all contributions. This gives the WITHIN element

Group R	
Pop share	0.429
Income share	0.220
Mean income	3,833
Covariance	355.6
Gini group R	**0.186**
Contribution to G(WIT)	**0.0175**
Group U	
Pop share	0.571
Income share	0.780
Mean income	10,200
Covariance	443.8
Gini group U	**0.087**
Contribution to G(WIT)	**0.0388**
Gini (WIT)	**0.056**

STEP 8, 9 and 10

Calculate subgroup mean incomes and replace actual incomes with the corresponding means. Calculate the Gini Index on this simulated income distribution

Fractional rank	Mean Incomes	Type (*)
0.286	3,833	R
0.286	3,833	R
0.286	3,833	R
0.786	10,200	U
0.786	10,200	U
0.786	10,200	U
0.786	10,200	U
Gini (BET)	**0.209**	

STEP 11

Calculate the residual K

Gini original distribution	**0.265**
Gini(BET) + Gini(WIT)	**0.265**
K	**0.000**

First of all, we must identify incomes belonging to individuals in different groups. The example is drawn by again considering rural and urban incomes. In Step 2, Table 2, incomes in each subgroup are ranked from the lowest to the richest. Note the meaning of **non-overlapping incomes**. When ranked from the lowest to the richest in the total income distribution (Step 1), each individual has exactly the same position as he/she has when incomes are separately ranked in each subgroup (Step 2). In other words, the final result of these two different ranking procedures is exactly the same. This very simple case, in the example, is obtained by assuming that the three poorest incomes belong to individuals living in rural areas. Extremely important, however, is the fractional rank assigned to this distribution. Each individual must be assigned the fractional rank calculated within each subgroup, i.e. instead of having a unique $F(y)$ we have two $F_i(y)$, with $i=R,U$ as if they were two distinct income distributions.

From Step 3 to Step 6, we must calculate a series of parameters. For **each group**, population shares, income shares and the Gini Index is actually calculated. For example, the rural group (R) represents 42.9 per cent of total population, it has 22 per cent of total income, and the Gini Index measured only on rural incomes is 0.186. Its contribution to **WITHIN** inequality is then obtained by multiplying the Gini Index by the population and income shares. It gives **0.0175**. The urban group (U) represents instead 57.1 per cent of the total population, it has 78 per cent of total income and the Gini Index measured only on urban incomes is 0.087. The contribution of this group to **WITHIN** inequality is therefore **0.0388**. It means that G_{WIT} is equal to 0.056 (0.0175+0.0388).

Steps 8 and 9 are now familiar. The procedure is the same as in the case of variance. Mean incomes replace actual incomes in each subgroup. The Gini Index calculated on this simulated income distribution is 0.209. This is G_{BET}, i.e. the **BETWEEN** element.

Now, in Step 10 we first calculate the Gini Index of the original income distribution of Step 1. This Gini is 0.265. Then, we can sum G_{WIT} and G_{BET}, which is again equal to 0.265. This means K=0, i.e. no residual. The Gini Index is therefore perfectly decomposable in the case where the distribution in Step 1 and the distribution in Step 2 gives rise to the same ranking. In this case, distributions are said to be **non-overlapping**.

Consider now Table 3, when the example of Table 2 is replicated assuming that rural and urban incomes overlap. This is clearly seen by comparing the income distribution in Step 1 and the corresponding distribution in Step 2, which give rise to different outcomes. Incomes are not ranked from the overall poorest to the overall richest, but from the poorest to the richest within each group. Indeed, one of the individuals living in rural areas comes from the top of the income distribution.

Now, the contribution of each group to G_{WIT} is **0.0492** and **0.0680** for rural and urban incomes, respectively. The **WITHIN** element is therefore equal to **0.117**. The **BETWEEN** element is calculated as before; it is now equal to **0.081**. Summing G_{WIT} and G_{BET} now gives **0.198**. The Gini Index of the original income distribution is, as before, **0.265**. Therefore, the difference between the two is K = **0.067**. The residual term is now positive, as the rank by subgroup incomes overlap with the rank of the total income distribution.

Table 3: Decomposing the Gini Index with residual

STEP 1

From the original income distribution, identify incomes belonging to different groups

Cumulative distribution function	Incomes	Type (*)
0.143	2,000	R
0.286	4,300	U
0.429	5,200	R
0.571	8,500	U
0.714	8,800	U
0.857	11,000	R
1.000	12,500	U
Total income	52,300	
Mean income	7,471	

STEP 2

Sort incomes in each subgroup

Subgroup cumulative distribution function	Incomes	Type (*)
0.3333	2,000	R
0.6667	5,200	R
1.0000	11,000	R
0.2500	4,300	U
0.5000	8,500	U
0.7500	8,800	U
1.0000	12,500	U

STEP 3, 4, 5, 6 and 7

Calculate population share and income share. Calculate subgroup Gini Indexes and multiply them by shares. This gives contributions to G(WIT). Sum all contributions. This gives the WITHIN element

Group R	
Pop share	0.429
Income share	0.348
Mean income	6,067
Covariance	1000.0
Gini group R	**0.330**
Contribution to G(WIT)	**0.0492**
Group U	
Pop share	0.571
Income share	0.652
Mean income	8,525
Covariance	778.1
Gini group U	**0.183**
Contribution to G(WIT)	**0.0680**
Gini (WIT)	**0.117**

STEP 8, 9 and 10

Calculate subgroup mean incomes and replace actual incomes with the corresponding means. Calculate the Gini Index on this simulated income distribution

Fractional rank	Mean Incomes	Type (*)
0.476	6,067	R
0.476	6,067	R
0.476	6,067	R
0.643	8,525	U
0.643	8,525	U
0.643	8,525	U
0.643	8,525	U
Gini (BET)		**0.081**

STEP 11

Calculate the residual K

Gini original distribution	**0.265**
Gini(BET) + Gini(WIT)	**0.198**
K	**0.067**

But what is the meaning of the term K? At this stage, a useful way to understand the meaning of K is to exploit the correspondence between the Gini Index and the Lorenz Curve. Let us use the example of Table 3 to draw Figure 4.

The **original income distribution** in Table 3 gives rise to the normal Lorenz Curve (the solid bold line in the graph); the between distribution gives rise to a Lorenz Curve which has a **kink** at the point where mean income changes (the solid line); the **within distribution** gives rise to the dotted concentration curve that touches the between concentration curves where the mean income changes. The bold dotted line is the equidistribution line.

Note that the within distribution gives rise to a **concentration curve** and **not to a Lorenz Curve**, as it is plotted against the overall cumulative distribution function and not against the within-group cumulative distribution function.

Figure 4: Gini Index decomposition, Lorenz Curves and residual

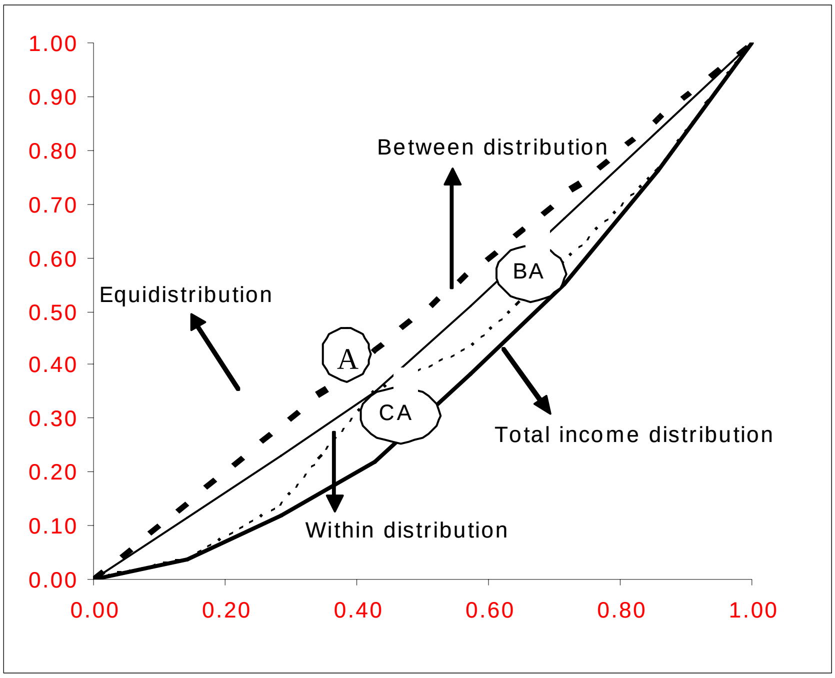


Figure 4 allows us to understand the three elements of the Gini coefficient in terms of areas:

- The difference between the **equidistribution line** (no inequality) and the Lorenz Curve of the **between** distribution is the inequality due to having different mean incomes (area **A**), i.e. the **between** element;

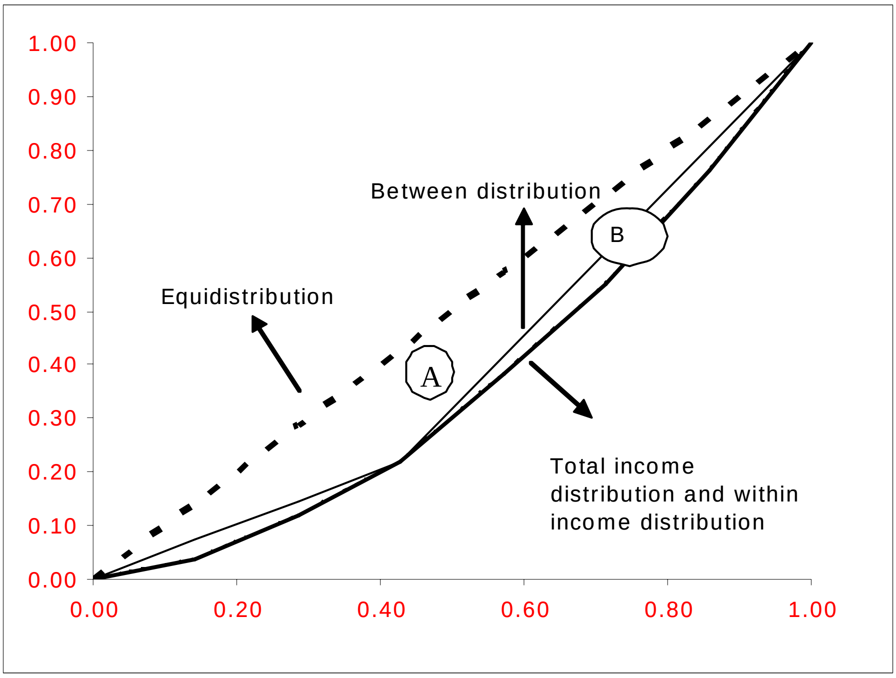
- The difference between the **between** distribution and the **within** distribution is the inequality due to having different actual incomes within each group (area **B**), i.e. the **within** element;

- The difference between the **within** curve and the normal **Lorenz Curve** is the **residual** term *K* (area **C**), i.e. the inequality due to the fact that the rank of the individual in the overall income distribution is not the same as its rank in the within-group income distribution. Individuals changing places when moving from the overall to the within-group distribution give rise to the **re-ranking effect**.

This interpretation suggests that were this re-ranking absent, the Gini Index would be perfectly decomposable in between and within elements, i.e. K=0. Graphically, re-ranking is absent when the within income distribution coincides with the overall income distribution. This occurs only when groups are NON-OVERLAPPING, i.e. the position of any given individual is the same both in the total income distribution and the within-group income distribution.

Figure 5 illustrates the case, drawing on the example reported in Table 2. Note that now the within distribution concentration curve and the Lorenz Curve of the total income distribution coincide. Area **C** is zero and K is zero.

Figure 5: Gini Index decomposition, Lorenz Curves and no residual



Summing up:

- The Gini Index is generally **not decomposable**, as it includes a residual term. This is true whenever groups are **overlapping**. In this case K conveys information on the rankings of individuals;
- The Gini Index is perfectly decomposable only in the special case where groups are non-overlapping;
- The interpretation of the residual term is not straightforward. This requires a careful interpretation of the economic meaning of the inequality decomposition.

5.3 An example of how to decompose the Theil Index

Table 4 reports an example of how to decompose the Theil Index. Steps 1 and 2 are, respectively, the original income distribution and the income distribution obtained by sorting incomes within each group. Note that the assumption is that rankings in total income distribution and in the subgroup income distribution overlap.

Table 4: Decomposing the Theil Index

STEP 1

From the original income distribution, identify incomes belonging to different groups

Individuals	Incomes	Type (*)
1	2,000	R
2	4,300	U
3	5,200	R
4	8,500	U
5	8,800	U
6	11,000	R
7	12,500	U
Total income	52,300	
Mean income	7,471	

(*) R=Rural; U=Urban

STEP 2

Sort incomes in each subgroup

Rank in original income distribution	Ordered incomes by subgroup	Type (*)
1	2,000	R
3	5,200	R
6	11,000	R
2	4,300	U
4	8,500	U
5	8,800	U
7	12,500	U

STEP 3, 4, 5 and 6

Calculate income shares. Calculate subgroup Theil Indexes and multiply them by shares. This gives contributions to Theil(WIT). Sum all contributions. This gives the WITHIN element

Group R	
Total income	18,200
Mean income	6,067
Income share	0.3480
Theil	**0.194**
Contribution to WITHIN	**0.067**
Group U	
Total income	34,100
Mean income	8,525
Income share	0.6520
Theil	**0.061**
Contribution to WITHIN	**0.040**
Theil (WIT)	**0.107**

STEP 7, 8 and 9

Calculate subgroup mean incomes and replace actual incomes with the corresponding means. Calculate the Theil Index of this simulated income distribution

Rank in original income distribution	Mean incomes	Type (*)
1	6,067	R
3	6,067	R
6	6,067	R
2	8,525	U
4	8,525	U
5	8,525	U
7	8,525	U
Theil (BET)	**0.014**	

STEP 10

Check the decomposition. The sum of Theil(WIT) and Theil(BET) must be equal to the Theil of the original income distribution

Theil original distribution	**0.121**
Theil (BET)+ Theil (WIT)	**0.121**

From Steps 3 to 6, we must calculate some parameters in order to derive the contribution of each group to the **within** element of inequality. The corresponding contribution to Theil Indexes are **0.067** for rural incomes and **0.040** for urban incomes. This gives Theil (WIT) = **0.107.**

In Steps 7 to 9, we must calculate the Theil Index of the simulated income distribution when the average income level of each group replaces the original income distribution. This is **0.014**. Summing Theil (BET) and Theil (WIT) gives the Theil of the original income distribution, i.e. **0.121** (Step 10). Note that the Theil Index, unlike the Gini Index, is perfectly decomposable even though incomes overlap (the income distributions in Steps 1 and 2 are different).

6. A SYNTHESIS

The most suitable inequality indexes for decomposition are the Gini Index and the generalised entropy class. Each of them has particular features. Table 5 below summarises what are their main characteristics and formulates a judgement on their relative usefulness for applied works.

Table 5 – Decomposability by subgroups

	Residual	Structure of weights	Appeal
GINI	Yes, in some cases	Do not sum up to one	Medium
GE_0	No	Sum up to one	High
T	No	Sum up to one	High
GEa	No	Do not sum up to one	Medium

The Gini Index may combine imperfect decomposability (with overlapping groups) and a structure of weights not summing to one. While it is a very good index for measuring inequality, its appeal for decomposability is at medium level. The same is true for members of the GE class with $\alpha>1$. Even though they are perfectly decomposable, weights do not sum up to 1. The best candidates for decomposability are the Theil Index and the mean logarithmic deviation, obtained by setting $\alpha=1$ and $\alpha=0$, respectively. They combine perfect decomposability with a nice structure of weights. This also explain their wide use in empirical applications.

7. READERS' NOTES

7.1 Time requirements

Time required to deliver this module is estimated at about four hours

7.2 EASYPol links

Selected EASYPol modules may be used to strengthen readers' background knowledge and to further expand their knowledge on inequality and inequality measurement.

This module belongs to a set of modules which discuss how to compare, on inequality grounds, alternative income distributions generated by different policy options. It is part of the modules composing a training path addressing **Analysis and monitoring of socio-economic impacts of policies.**

The following EASYPol modules form a set of materials logically preceding the current module, which can be used to strengthen users' background knowledge:

- ✓ EASYPol Module 000: ***Charting Income Inequality: The Lorenz Curve***
- ✓ EASYPol Module 001: ***Social Welfare Analysis of Income Distribution: Ranking Income Distribution with Lorenz Curves***
- ✓ EASYPol Module 040: ***Inequality Analysis: The Gini Index***

- ✓ EASYPol Module 051: ***Policy Impacts on Inequality: The Theil Index and the Other Entropy Class Inequality Indexes***

7.3 Frequently asked questions

- ✓ How to decompose inequality indexes?
- ✓ How to work out whether some groups of population contribute more to total inequality than others?
- ✓ Is decomposability by subgroups perfect?

8. REFERENCES AND FURTHER READING

Champernowne D.G., Cowell F.A., 1998. Economic Inequality and Income Distribution, Cambridge University Press, Cambridge, UK.

Cowell F.A., 1977. *Measuring Inequality*, Phillip Allan, Oxford, UK.

Gini C., 1912. *Variabilità e Mutabilità*, Bologna, Italy

Lambert P.J., Aronson J.R., 1993. Inequality Decomposition Analysis and the Gini Coefficient Revisited, *Economic Journal*, **103**, 1221-1227.

Seidl C., 1988. Poverty Measurement: A Survey, in Bös D., Rose M., Seidl C. (eds), *Welfare and Efficiency in Public Economics*, Springer-Verlag Heidelberg,, Germany.

Shorrocks A.F., 1980. The Class of Additively Decomposable Inequality Measures, *Econometrica*, **48**, 613-625.

Module metadata

1. EASYPol Module	052
2. Title in original language	
English	Policy Impacts on Inequality
French	
Spanish	
Other language	
3. Subtitle in original language	
English	Decomposition of Income Inequality by Subgroups
French	
Spanish	
Other language	

4. Summary

This tool illustrates how to decompose inequality measures by subgroups of populations. In particular, it defines the concepts of within and between inequality and analyses how different inequality indexes perform with respect to this decomposition. In particular, the performance of the analysis of variance, the Gini Index and the Theil Index will be discussed. A step-by-step procedure and numerical examples give operational content to the tool.

5. Date

December 2006

6. Author(s)

Lorenzo Giovanni Bellù, Agricultural Policy Support Service, Policy Assistance Division, FAO, Rome, Italy

Paolo Liberati, University of Urbino "Carlo Bo", Institute of Economics, Urbino, Italy

7. Module type	☐ Thematic overview ☐ Conceptual and technical materials ☒ Analytical tools ☐ Applied materials ☐ Complementary resources
8. Topic covered by the module	☒ Agriculture in the macroeconomic context ☐ Agricultural and sub-sectoral policies ☐ Agro-industry and food chain policies ☐ Environment and sustainability ☐ Institutional and organizational development ☐ Investment planning and policies ☒ Poverty and food security ☐ Regional integration and international trade ☐ Rural Development
9. Subtopics covered by the module	
10. Training path	[Analysis and monitoring of socio-economic impacts of policies]
11. Keywords	capacity building, agriculture, agricultural policies, agricultural development, development policies, policy analysis, policy impact anlaysis, poverty, poor, food security, analytical tool, income inequality, income distribution, income ranking, welfare measures, gini index, theil index, decomposition of income inequality, decomposition of indexes, entropy class indexes, social welfare functions, social welfare, variance analysis, structure of inequality