Artha Vijnana
Vol. LXII, No. 3, September 2020, pp. 219-224

A Simple Working Formula of Gini Coefficient with Some Common Software Command Codes

Amlan Majumder and Takayoshi Kusago

Although there has been a revolution in the computational methodology of Gini coefficient towards simpler formulations and easy computation procedure with the application of software, this paper presents a far simpler working formula of the same and consequently some software command syntaxes. The key component of the formula is now just a product of income-rank-based weight and income. The method will enable researchers, who work with a large number of observations, to compute the index and add it as a variable in the same working data file with ease.

I Introduction

There has been a revolution in computational methodology of Gini coefficient with the application of statistical software. The quest for a simple working formula and easy computation procedure has been the main driving force behind such endeavours. Given this background, although it seems saturated, the present paper still finds a scope to proceed further in this direction, particularly to work with the software and spreadsheet programmes for which ready-made command syntaxes for calculating Gini coefficient are not available. For example, the IBM SPSS Statistics, which is popularly (and previously) known as Statistical Package for the Social Sciences (SPSS henceforth) so far did not include any simple formula or single command in its menu for computation of Gini coefficient. However, the demand for the same among common users and academic researchers across disciplines is high. Sometimes researchers (who work with micro-data with a large number of observations) want to compute Gini coefficient and add it as a variable in the same working data file. A simple Google search on the subject matter will reveal lots of discussions and suggestions (on different possibilities of command syntax), many of which seem unworthy, often complex and at times confusing. Some of them appear to be trustworthy, but one may need adequate programming knowledge to follow them.[1]

Amlan Majumder, Department of Economics, University of North Bengal, Raja Rammohunpur, DT. Darjeeling 734013, West Bengal, Email: amlan@amlan.co.in

Takayoshi Kusago, Faculty of Sociology, Kansai University, 3-3-35, Yamate-cho, Suita-shi, Osaka 564-8680, Japan, Email: tkusago@kansai-u.ac.jp

We are thankful to anonymous referee for comments and suggestions. Responsibility of any error rests with us.

Luxembourg Income Study (LIS) offers computation of Gini coefficient for LIS data using SPSS through the online interface to registered users.[2] However, in this case, SPSS command syntax is not visible to users. The online interface of the World Bank also allows any of its user (even without any registration) to compute Gini coefficient using one's own data through Povcal Net Software.[3] However, availability of such facilities does not bridge the gap that arises from unavailability of simple SPSS command syntax for computing Gini coefficient and to have hand-to-hand experience on it. The same conclusion is true for common spreadsheet programmes too, such as Microsoft Excel or Apache Open Office Calc, etc. Availability of simple command could make these easily available spreadsheet programmes popular in computation of Gini coefficient, up to now the use of which for the said purpose is too low. In such a situation, this paper presents a simple working formula of Gini coefficient and consequent SPSS command syntax to compute the same and add it as a variable in the same working data file. Similar command syntax is also demonstrated for the said common spreadsheet programmes. The command syntaxes are meant for all the versions of the software packages. Special attention is given to 'simplicity' so that users with minimum programming knowledge can compute Gini coefficient from micro-data (with large number of observations) or else with ease.

II A Simple Working Formula for Computation of Gini Coefficient

Researchers in the field of measurement of economic inequality paid enough attention so far towards simple derivation of Gini coefficient. However, simplicity in formulation (working formula particularly) was probably a priority in an era when the use of computers or software did not overweigh the use of hand calculators. The classical example that we may cite in support of our understanding is the work of Milanovic (1997, pp. 45-46). He started with a "popular and easy formula" of Gini coefficient under the covariance approach[4] and from there he derived a further simple one.[5] In regard to the former, he noted that "STATA software calculates Gini in the same way". And for his own formula, he claimed that since all the "elements are easy to calculate, the Gini can be obtained using a simple hand calculator". However, even if we consider such quests for simplicity (from the point of view of authors) as a continuous process that is independent of context and time, we must realise that the demand of this era (from the point of view of users) sounds for more user-friendly software command, not for simple formula of Gini coefficient. Of course, in order to address the former, we need to pay attention to the latter.

Although there are various approaches to define Gini coefficient, we begin with the classical definition under the Gini's mean difference approach (*see* Anand, 1983, p. 313):

$$G_1 = (1/2n^2\mu)\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j|, \qquad \ldots(1)$$

where, G stands for Gini coefficient, the subscript denotes its sequence in the paper and y_i is the income of person or group i, y_j is that of person or group j, μ is the average income, i = 1, 2, 3, …, n, j = 1, 2, 3, …, n and $y_1 \leq y_2 \leq \ldots \leq y_n$. It means that i also stands for income-rank-based weights (in ascending order). For operational advantage, we plan to work with distribution of income. In that case, we comprehend that each of y_i and y_j is proportion or share of income corresponding to person or group i and j respectively (such that, $\mu = 1/n$, as $\Sigma y_i = 1$).

The final term in the right-hand side expression of the above formula measures income difference between two individuals. It is summed to compute total difference of income across the length and breadth considering all possible pairs of individuals. It is easily understood that, when there are n number of observations, the total number of pairs will be n^2 (n-square). Mean difference is computed by dividing the total difference of income by n^2 (n-square). It is then standardised with a division by μ, the mean income. Finally, half of the standardised mean difference is considered to define the Gini coefficient as per definition. Under a Lorenz curve framework, it is nothing but twice the area covered by the Lorenz curve and the egalitarian line.

Although the above method is classic in its appeal and noble in its approach to understand economic inequality, it does not make computation simple for large observations. For example, when n = 100, one needs to compute n^2 (n-square) or 10000 pairs of difference of income (or nearly half of it, as discussed below) using any software or spreadsheet programme. In order to minimise such physical workload, we proceed to derive a simpler form of the formula, as shown below.

The pairs of income difference of formula (1) may be presented in a symmetric matrix form. While working with it, Anand (1983, pp. 313-314) restricts the number of elements of it to the lower triangular portion of it for i = 1, 2, 3, …, n and $j \leq i$:

$$\begin{bmatrix} |y_1 - y_1| & |y_1 - y_2| \cdots & |y_1 - y_n| \\ |y_2 - y_1| & |y_2 - y_2| \cdots & |y_2 - y_n| \\ \cdots & \cdots & \cdots \\ |y_n - y_1| & |y_n - y_2| \cdots & |y_n - y_n| \end{bmatrix}. \qquad \ldots(2)$$

Gini index, in response to the above restriction, is expressed as following:

$$G = (1/n^2\mu) \sum_{i=1}^{n} \sum_{j \leq i} |y_i - y_j|, \qquad \ldots(3)$$

as sum of all (absolute) terms is twice the sum of the terms in the lower triangular matrix:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j| = 2 \sum_{i=1}^{n} \sum_{j \leq i} |y_i - y_j|. \qquad \ldots(4)$$

As,

$$\sum_{i=1}^{n}\sum_{j\le i}|y_i - y_j| = \sum_{i=1}^{n}\{(i-1)y_i - (n-i)y_i\}, \quad \ldots(5)$$

Equation (3) may be presented as:

$$G = \frac{1}{n^2\mu}\left(\sum_{i=1}^{n} iy_i - \sum_{i=1}^{n} y_i - n\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} iy_i\right). \quad \ldots(6)$$

After simplification, it reduces to:

$$G_2 = \frac{1}{n^2\mu}\left(2\sum_{i=1}^{n} iy_i - n - 1\right), \quad \ldots(7)$$

as, $\Sigma y_i = 1$.

As the composite term (product of income-rank-based weights and income) in the numerator of the above expression can be computed easily, it appears to be a very simple working formula of Gini coefficient. Further, the physical workload is now almost independent of the number of observations. For example, using SPSS or Microsoft Excel or Apache Open Office Spreadsheet, the composite term may be computed by a single command:

$$= i * y. \quad \ldots(8)$$

As by definition, $\mu = 1/n$, one does not need to write any programme to compute mean income.

Researchers in the field of measurement of economic inequality are well aware that when the number of observations is small, in order to avoid underestimation, the denominator n^2 (n-square) in formula (7) may be replaced by n(n-1) directly (see Deaton, 1997, p. 139). After such a replacement in formula (7), we get,

$$G_3 = \frac{1}{n(n-1)\mu}\left(2\sum_{i=1}^{n} iy_i - n - 1\right). \quad \ldots(9)$$

SPSS command syntax and that for any spreadsheet programme (Microsoft Excel or Apache Open Office Calc) in the next section will be based on formula (9) only, as it works equally well when the number of observations is small or large.

III Simple Command Syntax for Computation of Gini Coefficient and Its Application

Computation of economic inequality starts with distribution. Let us consider a distribution with n = 1000. Let us also define the variable of income distribution

as y. One needs to enter data under this variable (y) in the ascending order. After making income distribution ready, the following SPSS command syntax, as per formula (9), may be used to compute Gini coefficient.

Table 1. Simple SPSS Command Syntax for Computing Gini Coefficient

Line number	Command Syntax	Remarks
1	compute i = $casenum.	Creates income-rank-based weight.
2	compute n=1000.	One needs to define n.
3	compute μ = 1/n.	Computes mean of the distribution.
4	compute iy= i*y.	Computes the product term.
5	if (y ge 0) break = 1.	Creates one break variable.
6	aggregate	Computes Σiy.
7	/outfile=*	
8	mode=addvariables	
9	/break=break	
10	/Σiy = sum(iy).	
11	compute G = 1/(n*(n-1)*μ)*(2*Σiy-n-1).	Computes Gini index in the data file.
12	format G (f8.4).	Formats up to four decimal.
13	execute.	Execution command.

Note: One needs to put the value of n in the second line before executing the command.
Source: Self-elaboration.

The command syntax, which is presented in Table 1, may be written in an SPSS syntax file and after selection of the whole, it may be run (by a right-click on the mouse to select 'Run Current') to compute Gini coefficient from micro-data as well as from grouped-data (in a column). If we consider the distribution of first 1000 natural numbers and execute the above, we will get a Gini coefficient which is known to be and equal to 0.3333.

The paper uses hypothetical data for illustration. For example, Gini coefficient is computed for natural numbers with n = 1000. It describes a data series (in column): 1, 2, 3, …., 998, 999, 1000. Such a data series can easily be generated in common spreadsheet programmes (or by the SPSS command that appears in the first line of table 1, when the working file is defined). From this data series, one needs to compute a distribution first (i.e., variable y) and then proceed to the command syntax, as shown in Table 1.

One may check the result of this exercise with other available software or with other available command syntax of SPSS (as mentioned in the introductory section). For example, one may get the same result from PovcalNet. One should note that the latter needs data in two series: population distribution and income distribution, and the formula is equivalent to G_2.

In case of spreadsheet programmes, one may compute the composite term of the formula (9) by a simple command: =i*y and compute the column sum. The

Gini coefficient may then be computed putting all values in formula (9). The result of this exercise will exactly be the same as mentioned above (0.3333).

If someone is interested to work with formula (7), in case of large observations, the command syntax in the 11th line may be changed as: G = 1/(n*n*μ)*(2*Σiy-n-1).

IV Conclusion

Economists in the field of measurement of economic inequality have always been in the quest of presenting alternative and simpler ways to compute Gini coefficient. Such quests seem almost independent of context and time. In line with the same spirit, even if we have witnessed a revolution in computational methodology of the index with the application of statistical software in recent past, in this paper we try to present a far simpler working formula with some software codes. The presented formulae and consequent command syntaxes of the software can be used to compute Gini coefficient from micro-data (with a large number of observations) or grouped-data with ease.

Endnotes

1. We would like to cite Raynald's SPSS Tools at: http://www.spsstools.net/en/syntax/syntax-index/tests-of-inequality/many-tests-of-inequality-v5/ (accessed on 6th March 2019).
2. https://www.lisdatacenter.org/ (accessed on 9th March 2019)
3. http://iresearch.worldbank.org/PovcalNet/PovCalculator.aspx (accessed on 9th March 2019).
4. *See* Anand (1983), p. 315.
5. His previous effort, with similar objectives, appeared as Milanovic (1994), where he considered the vertical distance between the Lorenz curve and 45 degree line.

References

Anand, S. (1983), *Inequality and Poverty in Malaysia: Measurement and Decomposition,* New York: Oxford University Press.

Deaton, A. (1997), *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*, Baltimore: John Hopkins University Press.

Milanovic, B. (1994), The Gini-type Functions: An Alternative Derivation, *Bulletin of Economic Research*, 46(1): 81-90.

---------- (1997), A Simple Way to Calculate the Gini Coefficient, and Some Implications, *Economics Letters,* 56(1): 45-49.