Gini Coefficient Supplemental Handout

I. Conceptual Purpose and Definition of a Gini Coefficient

The Gini Coefficient is derived from country income or expenditure data and serves as an indicator of the distribution of wealth across a population. Consequently, the Gini Coefficient is a great tool for measuring and monitoring inequality within a population. As inequality has continued to rise globally, calculating the Gini Coefficient is an important tool in poverty statistics (OECD, 2008).

Definition

The Gini Coefficient is defined as the measure of income inequality on a 0 to 1 scale, with 0 representing perfect equality (everyone has the same income) and 1 perfect inequality (one person has all the money) (OECD, 2015).

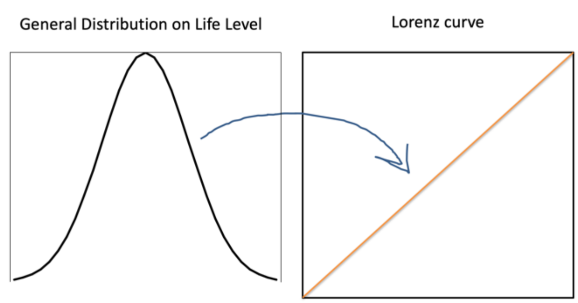
II. Fundamental Concepts

Lorenz Curve

As explained in the Poverty Statistics Module, the Lorenz Curve describes the cumulative proportion of income or expenditure plotted against the cumulative proportion of persons ranked according to income or expenditure.

When plotting the general distribution, inequality will appear as a skewed curve. This degree of inequality will thus be reflected by the profoundness of the Lorenz curve. These concepts are illustrated in Figures 1-3 below.

Figure 1. General income distribution plotted as a Lorenz curve.



The Lorenz Curve plots the data on cumulative income/expenditure and cumulative population proportions, and is thus a graphical representation of income inequality. In a world with perfect equality the Lorenz curve would exhibit a perfect slope of 1 at exactly 45 degrees, as seen in Figure 1. Meaningfully, this translates as 20% of the population having 20% of income, 50% having 50% of the income, 74% has 74%, and so forth.

In the real world, the Lorenz curve may appear more like Figure 2, or even Figure 3. The profoundness of the curve depends on the amount of inequality present in the population and appears steeper with increasing inequality, as observed in Figure 3. This may translate to 20% of the population having only 7% of the income, or maybe in extreme cases of poverty 65% of the population having only 38% of the income.

Figure 2. Asymmetric distribution of general income/expenditure plotted Lorenz curve.

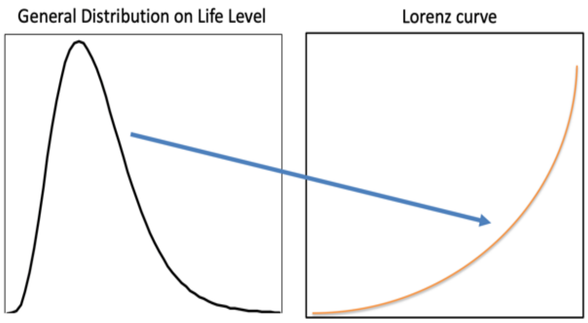
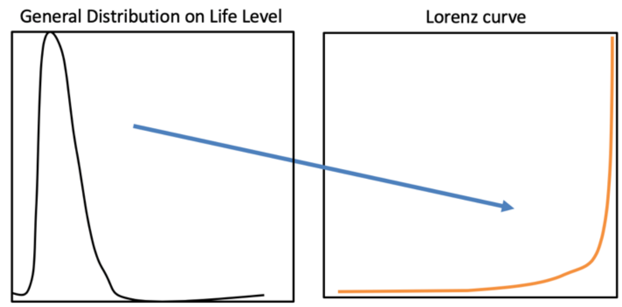


Figure 3. Increased skewness of Lorenz curve.

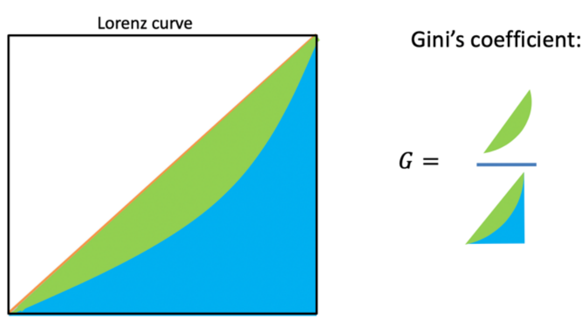


Gini Coefficient

The Gini Coefficient is an indicator utilized to assess the income distribution in a population based on the Lorenz curve (Bellú & Liberati, 2006). The measurement enables us to measure the degree to which the population strays from perfect equality. As described by Bellú and Liberati, perfect equality can be described as the linear equidistribution Lorenz curve depicted in Figure 1 (2006). We can thus assess the degree to which a society strays from this ideal measure by measuring the area between the equidistribution line and *our* plotted Lorenz Curve (2006). We are thus interested in the area within the perfect straight line and curve, also described as the concentration area, and highlighted in green throughout this supplementary handout (Bellú & Liberati, 2006).

As illustrated in Figure 4, the proportion of this area to the entire population represents the Gini Coefficient and can thus be used as an indicator of income inequality (Bellú & Liberati, 2006). As inequality grows and most of the income is concentrated among only a few people, the green concentration area becomes larger, thus the Gini Coefficient increases. In a world where only one individual has all the income, it is clear to see how the concentration area would encompass the entire right triangle and become 1, the maximum area (Bellú & Liberati, 2006).

Figure 4. Gini Coefficient Estimate



III. Calculating the Gini Coefficient

Data Considerations

To calculate the Gini Coefficient, either income or expenditure data may be used. Important considerations in utilizing expenditure-based datasets include (OECD, 2015):

- Different definitions of consumption and the collection of consumption data
- Gini estimates based on expenditure data often underestimates inequality

In utilizing income datasets, it is important to consider the following (OECD, 2015):

- Definition varies widely across data sets and countries
- Economies of scale (equivalency scales) must be taken into consideration
- When based on gross income, inequality is generally over estimated

Analysis Considerations

To be prepared for analysis, consider the following:

- How will set income brackets if the dataset has not predetermined income groups?
- For each bracket, is there a descriptive measure (mean, median) that can be utilized to plot the Lorenz curve?
- Ensure the dataset will enable you to calculate cumulative estimates of income and population distributions

Estimating the Gini Coefficient Utilizing Basic Geometry Concepts

In order to estimate the Gini Coefficient, we must thus estimate the proportion of the concentration area (green) to the total area (triangle containing green and blue areas). The total area can be easily calculated utilizing basic geometry of a right triangle: [(base x height)/2]. Since we are easily able to estimate the *total* area, we can estimate the concentration of the area by subtracting the area under the curve from the total area. Thus, we must know how to calculate this area under the curve. The simplest way is to break up the area under the curve and estimate the area of each individual trapezoid. We will apply this methodology in the next section.

IV. Example: Calculating the Gini Coefficient for Japan

The data below was extracted from Table 3 of the 2018 Japan Family Income and Expenditure Survey. Data was segregated into ten preset income distribution brackets. The relevant data for our example have been summarized in Table 1.

With this dataset, we will estimate the area of the 10 segments under the curve. Figure 5 illustrates this concept. To conduct area estimates of each segment, data points of average income within each income bracket and the proportion of households were utilized.

Calculation Methodology utilizing Excel

Table 1. 2018 Income Distribution among Japanese Households.

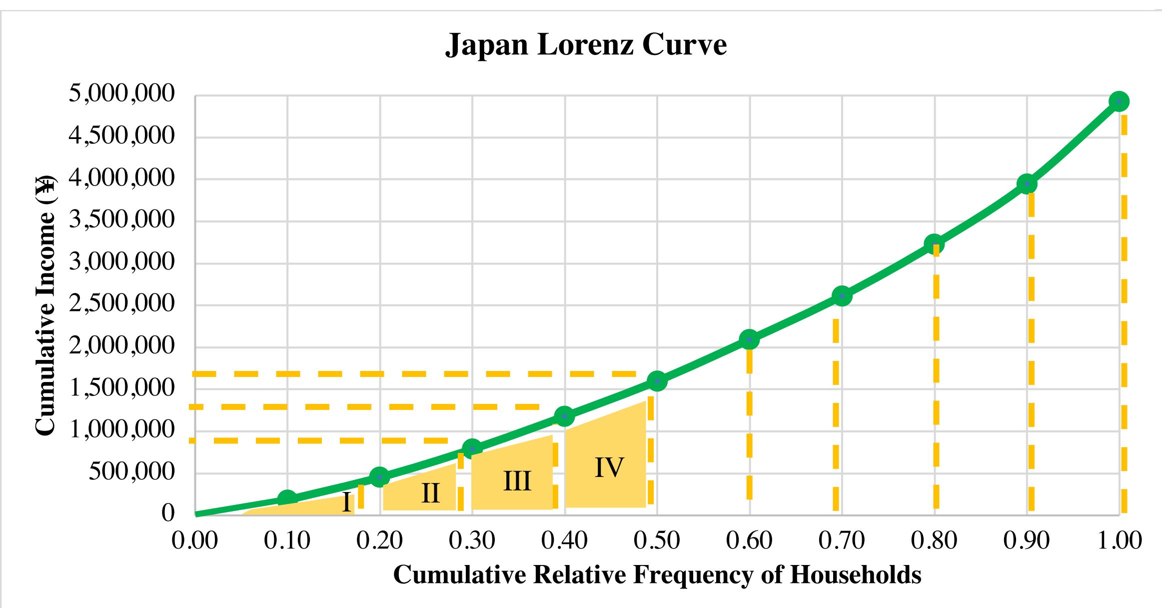
	Household		Income	
Income Group	**Relative Frequency**	**Cumulative Relative Frequency**	**Average**	**Cumulative Average**
I (< 2,580,000)	0.1	0.10	184,323	184,323
II (2,580,000-3,560,000)	0.1	0.20	269,141	453,464
III (3,560,000-4,290,000)	0.1	0.30	335,606	789,070
IV (4,290,000-4,980,000)	0.1	0.40	391,725	1,180,795
V (4,980,000-5,700,000)	0.1	0.50	420,062	1,600,857
VI (5,700,000-6,470,000)	0.1	0.60	492,457	2,093,314
VII (6,470,000-7,460,000)	0.1	0.70	520,786	2,614,100
VIII (7,460,000-8,610,000)	0.1	0.80	615,595	3,229,695
IX (8,610,000-10,480,000)	0.1	0.90	718,813	3,948,508
X (10,489,000 >)	0.1	1.00	977,435	4,925,943

Data Source: Table 3 Average of Monthly Receipts and Disbursements per Household by Yearly Income Quintile Groups, and by Income Decile Groups, 2018 Family Income and Expenditure Survey, Statistics Bureau of Japan.

First Step: Calculate cumulative relative frequencies of household population measure and income (or expenditure) averages.

→For visualization purposes, you may plot the relative population frequency and cumulative income average to derive the Lorenz Curve as seen in Figure 5.

Figure 5. Japan Lorenz Curve 2018.



Data Source: Table 3 Average of Monthly Receipts and Disbursements per Household by Yearly Income Quintile Groups, and by Income Decile Groups, 2018 Family Income and Expenditure Survey, Statistics Bureau of Japan.

Second step: Find the first area under the curve, which is the shape of a triangle so we will utilize the area of a triangle estimation (seen in Figure 6).

Third step: Find remaining area under the curve utilizing trapezoid formula for each section. Calculation details are outlined in Figure 6 below.

Figure 6: Calculations

First Area (Area of triangle under Group I):

$$= \frac{(184{,}323) \times (.1)}{2} \qquad = 9216.15$$

Second Area (Area of trapezoid under Group II):

$$= \frac{(453{,}464 + 789{,}070) \times (.1)}{2} \qquad = 62126.7$$

Third Area (Area of trapezoid under Group III):

$$= \frac{(789{,}070 + 1{,}180{,}795) \times (.1)}{2} \qquad = 98493.25$$

Repeat for remaining groups, estimates for each group are reported in Table 2.

Fourth Step: Sum all areas under the curve.

Fifth Step: Compute Gini Coefficient as follows:

$$= \frac{(Total\ Area) - (Concentration\ Area)}{Total\ Area}$$

$$= \frac{(2{,}462{,}971.5) - (2{,}092{,}790.75)}{2{,}462{,}971.5}$$

$$= .15$$

Table 2. Area Estimates.

Estimate of Gini	
Income Group	Area Under Lorenz Curve
I	9,216.15
II	62,126.70
III	98,493.25
IV	139,082.60
V	184,708.55
VI	235,370.70
VII	292,189.75
VIII	358,910.15
IX	443,722.55
X	246,297.15
Sum I-X	2,092,790.75
Total Area	2,462,971.50
Gini:	0.15

Once the area under the Lorenz Curve for each segment have been calculated, the areas must be summed. As we are able to estimate the entire area of the triangle (4,925,943 /2), we can subtract the concentration area from the total area to derive the concentration area and thus, the Gini Coefficient.

Useful Formulas

Area of a Triangle

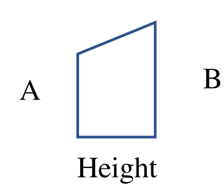
$$= \frac{Base\ x\ Height}{2}$$

Area of a Trapezoid

$$= \frac{(Side\ A + Side\ B)\ x\ Height}{2}$$

… equivlaent to…

$$= \frac{(Cum.Income_1 + Cum.Income_2)\ x\ (Relative\ Freq.)}{2}$$



Works Cited

Bellú, L. G. & Liberati, P. (2006). *Inequality analysis: the gini index*. Food and Agriculture Organization of the United Nations, FAO.

OECD Rights and Translation Unit. (2008). *Growing unequal?: income distribution and poverty in OECD countries*. OECD.

OECD. (2015). *In it together: why less inequality benefits all*. Paris, OECD Publishing. http://dx.doi.org/10.1787/9789264235120-en.