Measuring Income Inequality: an Exploratory Review

Region of Waterloo Public Health
Health Determinants, Planning and Evaluation Division
Alicja Krol and Judy Maan Miedema

June 2009

Introduction

More and more people are talking about the need to measure income inequality along side measurements of poverty. Poverty data explains what is going on at the bottom of the income ladder. Income inequality data looks at the distribution of income. Inequality implies that there are disparities in the distribution. Studies in the US suggest that level of income is an important determinant of absolute deprivation but that the level of income inequality in a community may be more important to understanding the implications for individuals and communities (FCM, 2003).

The exploratory review was done at the request of stakeholders in the Regional Municipality of Waterloo who wanted to understand the current methods being used to measure income inequality.

Measuring Income Inequality

Just as for measuring poverty, debates exist on how to measure income inequality in Canada (Raphael, 2007). Deciding on a standard measure is challenging because these measures are dependent on what society considers unequal. Choosing a standard inequality measure is really a choice between alternative definitions of inequality rather than a choice between alternative measures of a single theoretical construct (Allison, 1978).

Relative measures of income inequality compare the income of one individual group with the income of another group. These measures are most useful when analyzing the scope and distribution of income inequality. Examples presented in this paper include:

A. Median Share of income
B. Calculations based on percentile distributions
C. Lorenz Curve and the Gini Coefficient
D. Robin Hood Index
E. Atkinson Index
F. Theil's Entropy Measure
G. Coeficient of Variation

While all measures of income inequality have some limitations, if measures are calculated in a well explained and consistent way, they can provide a good tool for quantitative comparisons of inequalities over time.

A. Median Share of Income

Median share of income refers to the proportion of income held by households whose incomes fall below the median household income. It is a measure used by Statistics Canada and supported by the Association of Public Health Epidemiologists of Ontario.

The measure is calculated as follows:

1. Determine median household income (e.g. $20,000).
2. Add together the incomes of the bottom half of all households in area being studied (e.g. $1,000,000).
3. Add together all household incomes in the area being studied ($4,000,000).

4. Divide the total household income of the bottom half of households by the total household income of all households in that area ($1,000,000/$5,000,000).

Therefore the median share of income is 20%. A proportion of 50% would mean no inequality.

This measure is simple to calculate and uses readily available data. However, it is not sensitive to varying proportions of the income distribution within the upper or lower 50% of the distribution.

B. Calculations based on Percentile Distribution

There are several calculations of income inequality that can be done from tables which show the percentile distribution of income. Statistics Canada and the Canadian Centre for Policy Alternatives use measures described in this section.

To create a table with percentile distributions:

1. Obtain income data for every household in the geographic area you are studying in a given time period.
2. Then arrange the data in a list from the household with the highest income to the household with the lowest.
3. Divide the data into either equal fifths (quintiles) or equal tenths (deciles). A table showing this type of data might look like Table 1.

Table 1: Average Household Income by Decile (‘000), Waterloo Region, after Tax/Transfers, 1992 and 1996

	10%_ile	20%_ile	30%_ile	40%_ile	50%_ile	60%_ile	70%_ile	80%_ile	90%-ile
1992	xx	xx	xx	xx	xx	xx	xx	xx	xx
1996	xx	xx	xx	xx	xx	xx	xx	xx	xx

From this data you can calculate **percentage change over time** and add other regions for comparison.

Table 2: Per cent Change in Average Household Income by Decile, Waterloo and Canada, 1992-1996

	10%_ile	20%-ile	30%_ile	40%_ile	50%_ile	60%_ile	70%_ile	80%_ile	90%-ile
Waterloo	-17.2	-7.1	-4.3	-2.0	-1.2	-0.3	0.7	1.4	2.2
Canada	-18.2	-9.3	-7.8	-5.8	-4.5	-3.0	-2.1	-1.2	0

(Source: Federation of Canadian Municipalities, 1999)

From this data you can also calculate the **proportion of the total income** for the geographical area of interest held by each decile or quintile. You can look at the proportion before and after taxes and transfers to see the effect of government interventions. You can also see the direction and amount of change over a certain time period. Table 3 shows these calculations for Canadian data from 1984 and 1994.

Table 3: Proportion of Total Household Income by Quintile, Canada, before Transfers and Taxes, 1994, and after Transfers and Taxes, 1984 and 1994 (in percent)

Quintile	Q1	Q2	Q3	Q4	Q5 (highest)
Year					
1984 before transfers and tax	2.2	10.9	18.2	25.5	43.2
1984 after transfers and tax	7.1	13.3	18.4	23.9	37.3
1994 before transfers and tax	2.0	10.2	17.9	25.9	44.1
1994 after transfers and tax	7.7	13.4	18.3	23.8	36.8
1984-94 change (after taxes and transfers)	*+0.6*	*+0.2*	*-0.1*	*-0.1*	*-0.5*
1994 change ("after taxes and transfers" minus "before taxes and transfers")	*+5.7*	*+3.2*	*+0.4*	*-2.1*	*-7.3*

Source: Income After Tax, Distribution by Size in Canada (Ottawa: Statistics Canada, 1986 and 1996).

Lastly, from this data you can calculate the **ratio** of the top and bottom quintiles or top and bottom deciles in each year to arrive at an index of inequality – which is a single number that summarizes how equal in terms of income distribution a geographic region was in a certain year or time frame. Table 4 shows the results of this calculation for the same data as Table 3.

Table 4: Quintile Ratios (Q5/Q1)

Year	Quintile ratio(Q5/Q1)
1984 before transfers and tax	19.6
1994 before transfers and tax	22.1
1984 after transfers and income tax	5.3
1994 after transfers and income tax	4.8
Change in per cent (before)	+12.8
Change in per cent (after)	-9.4

Source: How the Pie is Sliced: Measuring Income Inequality in Canada, Nelson Society

This measure is also simple to calculate, uses readily available data, and is easy to interpret. It allows for comparisons to be made over time (including direction and magnitude) and can also be used to calculate how effective the government was in redistributing income from year to year.

C. Gini Coefficient

The Gini coefficient is derived from the Lorenz curve illustrated in Figure 1. The Lorenz curve plots the cumulative share of income earned on the y-axis by households ranked from the bottom to the top on the x-axis. In a region with perfect equality, the Lorenz curve would be a perfectly straight 45° line. As inequality increases, the Lorenz curve deviates away from the line of perfect equality (Borooah et al, 1991).

The Gini coefficient is the area between the line of perfect equality and the observed Lorenz curve, as a percentage of the area between the line of perfect equality and the line of perfect inequality. The Gini coefficient varies between 0, which reflects total equality (i.e. where everyone has the same income) and 1, which reflects total inequality (i.e. where one person has all the income, and everyone else has zero income). This is a comparison measure and is used most effectively when comparisons of Gini coefficients for different regions or time periods are made. It is a measure used by Statistics Canada, The World Bank, and the United Nations.

To calculate the Lorenz Curve:

The following is an example taken from the Department of Mathematics website, Washington University.

1. Calculate the proportion of total income (Table 5) – as described in section B of this report.

Table 5: Proportion of Aggregate Income

	Lowest 20%	Next Lowest 20%	Middle 20%	Second Highest 20%	Highest 20%
1968	4.2	11.1	17.5	24.4	42.8

Source: *Modified from the Department of Mathematics Website, Washington University*

2. Compute the cumulative fraction of aggregate income (See Table 6)

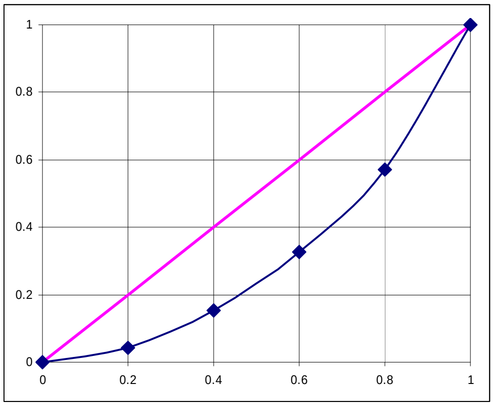
Table 6: Cumulative Fraction of Aggregate Income

	Lowest 20%	Lowest 40%	Lowest 60%	Lowest 80%	Lowest 100%
1968	.042	.153	.328	.572	1.00

Source: *Modified from the Department of Mathematics Website, Washington University*

3. Plot the six points on the Lorenz curve; (0, 0), (.2, .042), (.4, .153), (.6, .328), (.8, .572), (1, 1). (See Figure 2)

Figure 1:



Calculation of the Gini Coefficient:

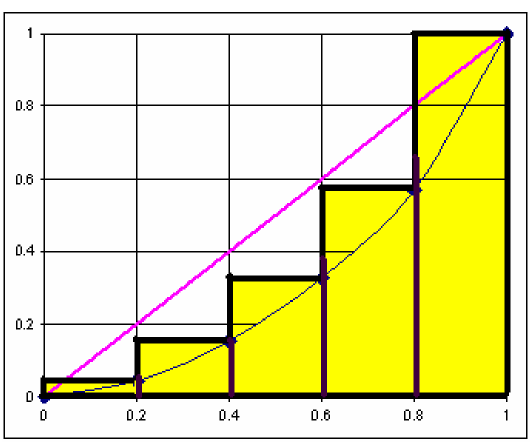
1. Find the area under the Lorenz curve between x=0 and x=1

Since the formula for the curve is unavailable due to few data points, numerical integration will be used.

The first method of integration is to compute the right sum. In each of the five vertical strips, the right most point of the curve is selected. A horizontal line is drawn from that point to the left edge of that vertical strip. This gives the horizontal lines of a staircase. The area under the staircase is the estimate of the area under the curve. The area under the staircase is found by multiplying the area of each base (=.2) by the height.

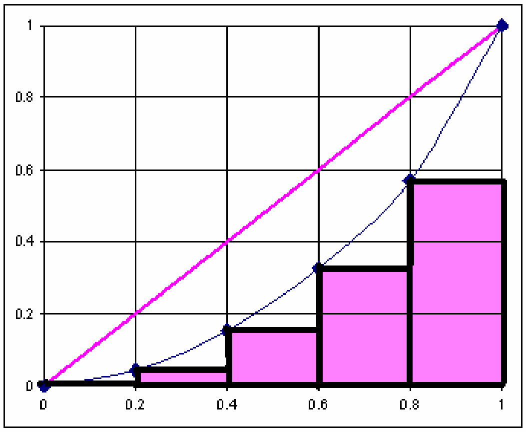
$(.2 \times .042) + (.2 \times .153) + (.2 \times .328) + (.2 \times .572) + (.2 \times 1.00) = .419$ (See Figure 3)

Figure 2:



The second method of computation is called the left sum. To compute this, follow the steps above using the leftmost curve instead of the right.
$(.2 \times 0) + (.2 \times .042) + (.2 \times .153) + (.2 \times .328) + (.2 \times .572) = .219$ (See Figure 4)

Figure 3:

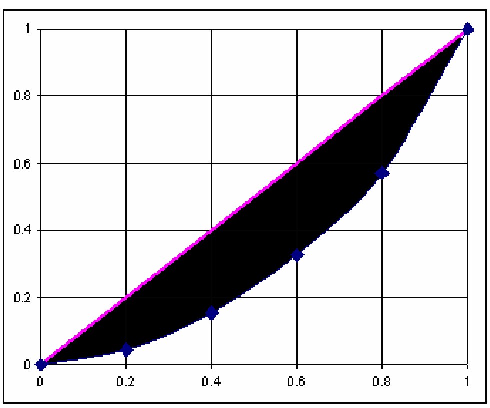


Although both of these approaches are fairly reasonable it is clear from the pictures that the right sum overestimates the area and the left sum underestimates the area. Hence a better estimate would result by averaging the two (.419 + .219)/2 = .319.

3. Compute the ratio between the diagonal line and the curve

The triangular area under the diagonal has size .5 (it is ½ of a 1 × 1 square). The shaded area is that triangular area of .5 minus the area under the curve which we just estimated as .319. Therefore, the estimate of the shaded area is .5 - .319 = .181. The ratio of this to .5 is .181/.5 = .362 which is the Gini coefficient.

Figure 4:



Source: *Modified from the Department of Mathematics Website, Washington University*

The Gini coefficient is useful as it can be calculated using either individual or household level data. Comparisons can be made over time or between different locations.

The Gini coefficient measures the size of an area rather than the shape so it is incapable of showing different kinds of inequality represented by various shapes of Lorenz curves. This means that economies with similar Gini coefficients can still have very different income distributions. As an extreme example, an economy where half the households have no income, and the other half share income equally has a Gini coefficient of ½; but an economy with complete income equality, except for one wealthy household that has half the total income, also has a Gini coefficient of ½ (DeMaio, 2007).

The Gini coefficient is most sensitive to inequalities and income transfers in the middle part of the income spectrum and does not emphasize inequalities in the top or bottom of the spectrum (polarization). For example, a Gini coefficient may indicate decreasing inequality but society may be becoming more polarized. The Gini simply indicates the spread of the income distribution or deviation from the mean. Other measures are better suited to show income polarization.

The Gini coefficient shows the direction of income redistribution but does not indicate where the redistributions are occurring (Moran, 2003). For example, the Gini coefficient may indicate declining inequality in a situation where income is redistributed only among the very poor while the overall structure of the income distribution remains the same.

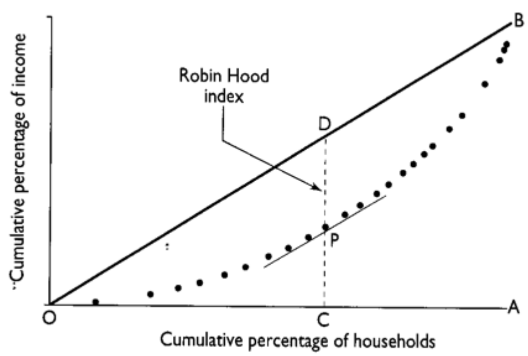
The Gini ignores life cycle effects – such as how in Western societies, an individual tends to start life with little or no income, gradually increase in income till about age 50, after which incomes will decline. This will have the effect of significantly overstating inequality (Wikipedia, 2008).

The last caveat of the Gini coefficient is that it does not allow for within or between income group comparisons.

D. Robin Hood Index

The Robin Hood index is related to the Lorenz curve and the Gini coefficient. It measures the portion of total income that would need to be distributed in order for there to be perfect equality. This measure is equivalent to the maximum vertical distance between the Lorenz curve and the line of perfect equality (45° line) as shown in Figure 6.

Figure 5:



The Robin Hood Index is easy to interpret and fairly easy to calculate once the Lorenz Curve has been calculated. However, much like the Gini, it not sensitive to income transfers between households on the same side of the mean income.

E. Atkinson Index

The Atkinson Index is an inequality measure that explicitly incorporates normative judgments about social welfare (De Maio, 2007). The value of the Atkinson Index can vary between 0 and 1. Like the Gini Coefficient, the Atkinson index is most effectively used in comparisons between regions. A lower Atkinson value represents an income distribution that is more equal. In addition, this measure incorporates a sensitivity parameter (ε) which can range from 0 to infinity. As the sensitivity index approaches higher values, the Atkinson Index becomes more sensitive to changes at the lowest income groups. As the sensitivity index approaches 0, the Atkinson Index becomes more sensitive to changes in the income position of the higher income groups in a distribution. It is common to see sensitivity values of 0.5, 1, 1.5 or 2.

This measure is calculated as follows:
1. Calculate the equity sensitive average income

$$y_e = \left(\sum_{i=1}^{n} f(y_i) \cdot y_i^{1-\epsilon} \right)^{1/(1-\epsilon)}$$

Where y_i is the proportion of total income earned by the *i*th group, and e is the sensitivity parameter.

2. Calculate the Atkinson Index

$$I = 1 - y_e / \mu$$

Where (*I*) represents the Atkinson Index and μ is the mean income. If an income distribution in a region is more equal, the value of y_e will be closer to μ. This will result in a lower value of the Atkinson Index.

The Atkinson index is valuable in that it incorporates a sensitivity parameter into the equation, but this sensitivity parameter involves a judgment about inequality. The index is also not very intuitive.

F. Theil's Entropy Measure

This measure is based on an income contribution or share that each individual or group holds. It involves complex mathematical calculations.

When individual data is available, each individual has an identical population share (1/N), so each individual's Theil's Entropy measure is determined by his or her proportional distance from the mean. When individual data is not available, the Theil's Entropy measure can be adjusted for groups. The index has a potential range from zero to infinity with higher values indicating more equal distribution of income.

Individual Level Calculation:

$$T = \sum_{p=1}^{n} \left\{ \left(\frac{1}{n} \right) * \left(\frac{y_p}{\mu_y} \right) * \ln \left(\frac{y_p}{\mu_y} \right) \right\}$$; the number of individuals in the population, y_p represents
ɔ and μ_y represents the population's average income.

Theil's measure is useful in that it adjusts for incomes at both the top and bottom of the income distribution.

# individuals	Income
2	$100, 000
4	$80, 000
6	$60, 000
4	$40, 000
2	$20, 000

(Modified from University of Texas Tutorial)

Individuals in the top salary group of this distribution contribute large positive elements to the Theil's Entropy measure. Individuals in the middle salary group contribute nothing to Theil's Entropy measure because their salaries are equal to the population average. Individuals in the bottom salary group contribute large negative elements.

Group Level Calculation:

Group data can be aggregated and used to calculate the Theil's Entropy measure. Members of the population are classified into mutually exclusive groups. The Theil's Entropy measure is then calculated based on two components, the between group and within group component.

T = T'g + Twg

(*T'g*) represents the between group component and (*Twg*) represents the within group component. When aggregated data is available instead of individual data, *T'g* can be used as a lower bound for Theil's measure in the population.

$$T'_g = \sum_{i=1}^{m} \left\{ \left(\frac{p_i}{P} \right) * \left(\frac{y_i}{\mu} \right) * \ln \left(\frac{y_i}{\mu} \right) \right\}$$

The variable *pi represents* the population of group *i*, *P* represents the total population, *yi* represents the average income in group *i*, and the variable *μ* represents the average income across the entire population.

# Individuals per group	Avg. group income
2	$95, 000
4	$75, 000
6	$60, 000
4	$45, 000
2	$25, 000

(Modified from University of Texas Tutorial)

The two top salary groups in this distribution contribute large positive elements to the Theil's Entropy measure. Middle salary groups contribute nothing to Theil's measure because their salaries are equal to the population average. The two groups in the bottom salary range contribute large negative elements.

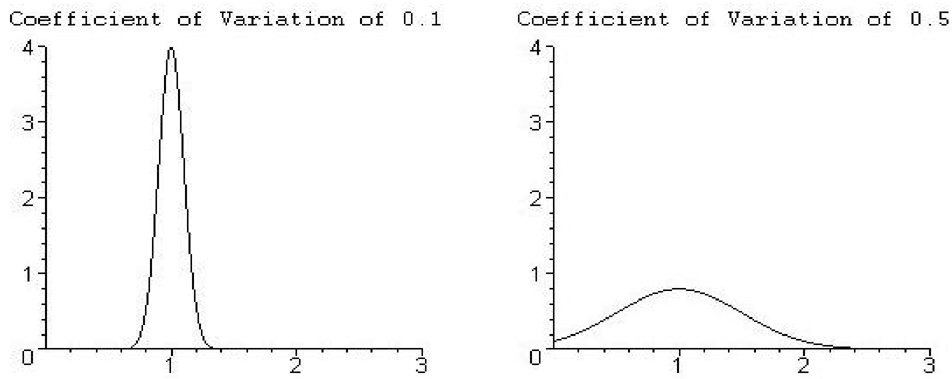
This measure is useful in that it allows the researcher to understand the contributions to inequality by within group and between group components. However, there are a number of caveats with Theil's Entropy measure. It is complex to calculate and hard to interpret. It varies greatly when the distribution varies regardless of whether the change in distribution occurs at the top, middle, or bottom. Income redistributions will impact the calculation irrespective of whether the redistribution takes place between the rich and the poor or the rich and the middle. Lastly, this measure cannot be used to directly compare populations with different sizes or group structures because the calculation is dependent on number of individuals in the population or group.

G. Coefficient of Variation (CV)

The coefficient of variation (CV) is a simple statistical method of representing the inequality of an income distribution. It is not commonly used.

To calculate the Coefficient of Variation:
Divide the standard deviation of an income distribution by the mean of the same distribution. Coefficients of Variation can be graphed as follows.

Figure 6:



(Modified from University of Texas Tutorial)

More equal income distributions will yield smaller CV values due to smaller standard deviations. For example, the graph on the left yields a smaller CV value because the standard deviation of the income distribution is smaller.

The coefficient of variation is simple to calculate but requires comprehensive individual data. Also, the mean and standard deviation used to calculate this measure are influenced by outliers such as high or low income values. Therefore, if income is not normally distributed, this measure would not be appropriate.

Conclusions

Table 7 summarizes the information found in this paper and makes some recommendations about the use of the various measures.

Table7: Summary of income inequality measures				
Measure	**Complexity of Calculation**	**Benefits**	**Caveats**	**Recommendation**
Median share of Income	easy	-data readily available	-not sensitive to varying proportions of the income distribution within the upper or lower 50% of the distribution	use in combination with other measures
Calculations based on Percentile distributions	easy	-data readily available -easy to interpret -allows for comparisons over time (including direction and magnitude) -used to calculate effectiveness of government transfers over time		use
Lorenz Curve and Gini Coefficient	complex but aided by statistical software	-a graphical representation of income inequality that can be compared over time and between geographic areas -simple to calculate -data readily available -can be calculated for individual and household level data -easily interpreted when compared to other Gini coefficients	-incapable of showing different kinds of inequality represented by various shapes of Lorenz curves -does not emphasize inequalities in the top or bottom of the spectrum (polarization) -shows the direction of income redistribution but does not indicate where the redistributions are occurring -ignores life cycle effects -does not allow for within or between income group comparisons	use
Robin Hood Index	easy if have Lorenz curve	-uses the same data needed to calculate the Lorenz curve -easy to interpret	-not sensitive to income transfers between households on the same side of the mean income	use together with the Gini coefficient
Atkinson Index	complex	-incorporates a sensitivity parameter directly into the equation.	-sensitivity parameter means that a subjective judgment has been made about inequality -not intuitive	do not use
Theil's Entropy Measure	complex	-shows the contributions to inequality by within group and between group components	-complex to calculate and interpret. -varies greatly when the distribution varies regardless of whether the change in distribution occurs at the top, middle or bottom -income redistributions will impact the calculation irrespective of whether the redistribution takes place between rich and poor or rich and middle -cannot directly compare populations with different sizes as calculation is dependent on number of individuals in the population or group	do not use
Coefficient of Variation	easy		-requires comprehensive individual data -not intuitive -cannot use if the income distribution is not normal	do not use

References

Borooah, VK, McGregor, PPL & McKee, PM (1991). Regional Income Inequality and Poverty in the United Kingdom. Dartmouth Publishing Company: Vermont.

DeMaio, F.G. (2007). Income inequality measures. *Journal of Epidemiology & Community Health,* 61(10) 849-852

Federation of Canadian Municipalities (1999). Quality of Life in Canadian Communities.

Federation of Canadian Municipalities. (2003). Falling behind: our growing income gap.

Moran, T. (2003). On the Theoretical and Methodological Context of Cross-National Inequality Data. International Sociology. 19(2):351-378.

Nelson Society. 2008. How the pie is sliced: measuring income inequality in Canada. Retrieved February 16, 2008 from http://www.nelson.com/nelson/harcourt/sociology/newsociety3e/research/research07.htm

Raphael, D. (2007). Poverty and Policy in Canada. Implications for Health and Quality of Life. Canadian Scholars' Press. Toronto, Ontario.

Statistics Canada (1996). Income after tax, distribution by size in Canada 1994. Statistics Canada, Ottawa.

University of Texas Tutorial. Measuring inequality: an examination of the purpose and techniques of inequality measurement. Retrieved February 10, 2008 from http://utip.gov.utexas.edu/tutorials/intro_ineq_studies.ppt

Washington University, Department of Mathematics, Course Notes, Math 128. Fall 2005. Retrieved February 20, 2008 from http://www.math.wustl.edu/~rr/128/gini.doc

Wikipedia. 2008. Gini Coefficient. Retrieved February 20, 2008 from http://en.wikipedia.org/wiki/Gini_coefficient