Using the Gini Coefficient and Other Measures of Inequality in a Quantitative Literacy Context

Dr. Michael Catalano, Dakota Wesleyan University
Michael.Catalano@dwu.edu
October 13th, 2018

Today, there is a huge disparity between the technology, education, health care and agricultural methods that are available in the developed and developing world. The principal challenge we face is to close that gap…The countries, businesses and individuals that are on the right side of the divide have to think hard about what kind of world they want us all to live in 20 years from now. Narrowing the gap benefits everyone, and we have the means to do it. If we don't, we will have missed an amazing opportunity. – Bill Gates (2000)

Questions to consider

- Why should we discuss social issues like inequality in a QL course?
- Numeracy involves not only correct mathematics, but making appropriate choices. What should we measure, why, and how?
  - There are often many ways to measure things, including inequality. What are the advantages and disadvantages of these various ways, especially with respect to a particular student audience?
  - Even if students do not do actual measurements, knowing how data is produced is relevant in using that data in reaching conclusion or making decisions.

Why include inequality in a QL course?

- Inequality is receiving an increasing amount of attention.
- Inequality is a concept most students can engage with on an intuitive level.
- Inequality can be measured in a number of different ways.
  - Instructors can choose measures appropriate to the given student audience, learning outcome, or content topic.
  - The examination or comparison of measures can be instructive.
- Inequality naturally involves both facts and judgments.
- Inequality is not likely to be a topic previously studied by students in a mathematics class.
- Inequality and other social justice issues are a natural way to promote the ideal of "civic virtue." (See William Briggs Numeracy perspective *Quantitative Literacy and Civic Virtue* from July 2018)

What is QL? – One partial view

In our 2009 Numeracy article, *Measuring Resource Inequality – The Gini Coefficient*, we specified the following five skills as part of what it means to be quantitatively literate.

1. The ability to understand quantitative information within a variety of real-world contexts, including the relevance and meaning of the given quantitative information in that context.
2. The ability to use mathematical and statistical methods to increase understanding, solve problems, and reach conclusions within a variety of real-world contexts. This would include being able to use techniques from different areas of mathematics within a single context.
3. The ability to critique the reasonableness of statements that include quantitative information or have quantitative information as support.
4. The ability to use and develop appropriate representations of quantitative information and use these within the context of written and oral communications.
5. The ability to understand mathematical and quantitative content from a conceptual standpoint, not simply as a set of algorithms or procedures.

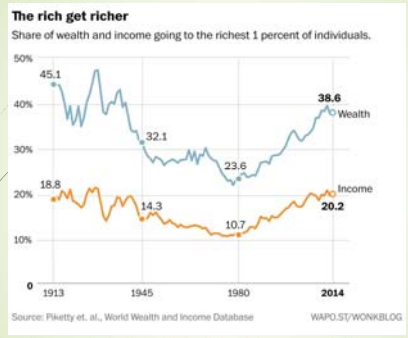
The Gini Coefficient is an appropriate topic to promote these goals, particularly 1, 4, and 5.

More things to think about

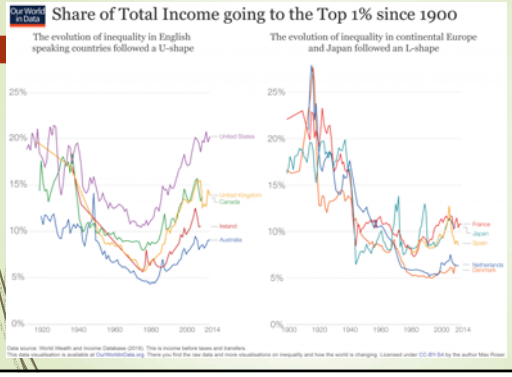
- How should we measure things, including inequality.
  - What makes a good measure?
    - Can values be easily related to examples? (Ht. vs. GDP)
    - Is it simple to calculate?
    - Is it intuitive to understand what is being measured and how?
    - Does it facilitate useful comparisons?
  - There are many ways to measure inequality. What are the advantages and disadvantages of these various ways?
    - What works in your class?

Measuring Inequality: Percentile shares

- What percentage or ratio of the resource is shared by those above a particular percentile, or between two percentiles?
- The graph here is for the U.S., from the Cato Institute, citing Piketty, *Capital in the Twenty-First Century.*
- Deciles and quintiles are common special cases of percentiles.
- Percentile shares are intuitive and fairly easily calculated from raw data.



https://www.washingtonpost.com/news/wonk/wp/2018/02/06/how-rising-inequality-hurts-everyone-even-the-rich/?utm_term=.f2a80baff2fc



https://ourworldindata.org/income-inequality

Questioning Quintiles: Implications of Choices of Measures for Income Inequality and Social Mobility by Joel Best

- Americans are said to prize the country's reputation for "social mobility."
- Using income data based on quintiles is a common way to measure social mobility.
  - Are average incomes within quintiles increasing?
  - Do individuals or their children move between quintiles? Can I make it to the top?
  - What counts as a 'socially mobile' society?
  - How is inequality between quintiles changing over time?
    - Is increasing inequality a bad thing, and if so, why?

Problems with choosing Quintiles as a Measure

- Using household data obscures differences in average household size and other household characteristics between quintiles.
- The total population of the top quintile in 2016 was 78.3 million, with only 44.5 million in the bottom quintile.
- Quintiles are zero sum. If a household goes in, one must come out.

Table 1
Characteristics of Bottom- and Top-Quintile Households, 2016 (thousands of households)

Characteristic	Bottom Quintile <$24,002	Top Quintile >$121,018
Households	25,245	25,245
Married couple families	4,388	19,314
Males living alone	5,583	1,158
Females living alone	8,817	678
Head >65	9,527	3,886
No earners	15,794	970
Two or more earners	1,090	18,773
Head employed full-time	4,533	19,619

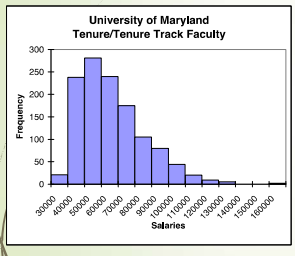
Table from Best (2018), page 3.

Best notes that one indicator inequality is increasing is that the upper middle class is getting larger, from 12.9% of the population in 1979 to 29.4% in 2014. The upper class went from 0.1% to 1.8%.

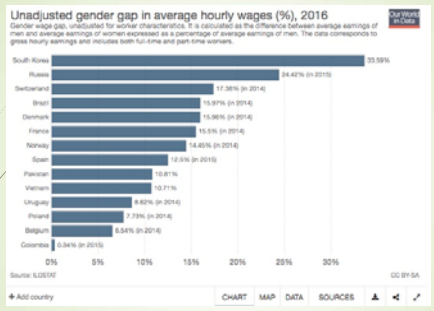
More factoids from Best

- Only 10% of bottom quintile college grads stay there, and 10% rose to the top.
- In a Brookings study, 51% of African-Americans stayed in the bottom quintile, and 3% made it to the top.
  - The comparable figures for White Americans were 23% and 16%.
- "About two-thirds of Americans wind up earning higher incomes than their parents, and the percentage of those outearning their parents is nearly twice as high for those whose parents were in the bottom quintile (82 percent) than those with top-quintile parents (43 percent)"
- Census Bureau statistics indicate that 43 percent of households moved—either up or down—into a new income quintile between 2009 and 2012.
  - Does this indicate we are a "socially mobile" society?
- *Certainly the facts and observations presented by Best can provide the grist for interesting questions or exercises one could post in a QL course.*
- *Given the readily available and understandable data, percentiles are a fairly easy way to bring a discussion of inequality into a QL course.*

Histograms



- Histograms are a natural way to represent distributions, and when the variable represented is a 'resource,' one can consider the histogram a representation of inequality.
- Bolker and Mast in their *Common Sense Mathematics* text discuss inequality briefly using distributions as a way to represent inequality.



https://ourworldindata.org/

Measuring Inequality: Ranges and Range Ratios

- Range: Manny Ramirez's salary minus Lenny Dinardo's is $22,200,000.
- Range Ratio: Divide instead of subtract to find that Manny made 75 times as much as Lenny.
- Advantages:
  - Intuitive, easy to calculate.
  - The ratio has no units.
- Disadvantages:
  - Ignores most of the data.
  - Skewed by outliers.
  - Ratio ignores absolute values.

Player	Salary	Player	Salary
Ramirez, Manny	22,500,000	Embree, Alan	3,000,000
Martinez, Pedro	17,500,000	Timlin, Mike	2,500,000
Schilling, Curt	12,000,000	Mueller, Bill	2,100,000
Garciaparra, Nomar	11,500,000	Reese, Pokey	1,000,000
Damon, Johnny	8,000,000	Mirabelli, Doug	825,000
Varitek, Jason	6,900,000	Burks, Ellis	750,000
Ortiz, David	4,587,500	Kapler, Gabe	750,000
Lowe, Derek	4,500,000	Daubach, Brian	500,000
Nixon, Trot	4,500,000	McCarty, David	500,000
Wakefield, Tim	4,350,000	Bellhorn, Mark	490,000
Mendoza, Ramiro	3,600,000	Arroyo, Bronson	332,500
Foulke, Keith	3,500,000	Crespo, Cesar	309,000
Kim, Byung-hyun	3,425,000	Shiell , Jason	303,000
Millar, Kevin	3,300,000	Garcia, Reynaldo	301,500
Williamson, Scott	3,175,000	DiNardo, Lenny	300,000

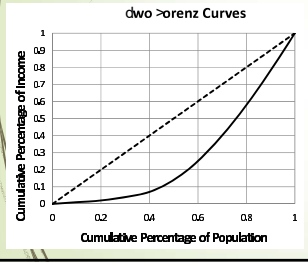
Measuring Inequality: Percentile Ranges and Ratios

- Interquartile range. Roughly half of the players make between $500,000 and $4.5 million (Daubach and Lowe)
- Percentile ratios: For example, take the 75th percentile divided by the 25th. Derek Lowe made 9 times as much as Brian Daubach.
- Not skewed by outliers.
- Intuitive and in common use.
- Easy to calculate.
- Still only based on two values.

Player	Salary	Player	Salary
Ramirez, Manny	22,500,000	Embree, Alan	3,000,000
Martinez, Pedro	17,500,000	Timlin, Mike	2,500,000
Schilling, Curt	12,000,000	Mueller, Bill	2,100,000
Garciaparra, Nomar	11,500,000	Reese, Pokey	1,000,000
Damon, Johnny	8,000,000	Mirabelli, Doug	825,000
Varitek, Jason	6,900,000	Burks, Ellis	750,000
Ortiz, David	4,587,500	Kapler, Gabe	750,000
Lowe, Derek	4,500,000	Daubach, Brian	500,000
Nixon, Trot	4,500,000	McCarty, David	500,000
Wakefield, Tim	4,350,000	Bellhorn, Mark	490,000
Mendoza, Ramiro	3,600,000	Arroyo, Bronson	332,500
Foulke, Keith	3,500,000	Crespo, Cesar	309,000
Kim, Byung-hyun	3,425,000	Shiell , Jason	303,000
Millar, Kevin	3,300,000	Garcia, Reynaldo	301,500
Williamson, Scott	3,175,000	DiNardo, Lenny	300,000

Lorenz Curves and the Pietra Index

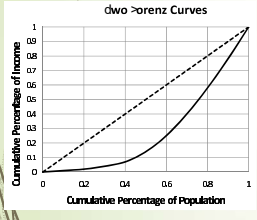
Cumulative ratio of pop.	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Cumulative ratio of resource	0.00	0.01	0.02	0.04	0.07	0.14	0.25	0.39	0.58	0.78	1.00



- A Lorenz curve plots cumulative ratio of a resource versus the cumulative ratio of the population. This graph can be considered a representation of inequality all by itself.
- If everyone in the population shares equally in the resource, the Lorenz curve is the line y = x.
- The Pietra Index is the vertical distance from the equality line and the Lorenz curve. This can be found by locating the coordinates of the intersection of the curve and the tangent line with slope 1.
  - The coordinates of this point are referred to as the mean resource and mean population share for the distribution.

Lorenz Curves and the Gini Coefficient

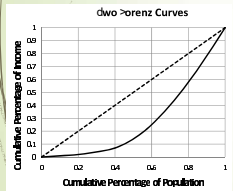
Cumulative ratio of pop.	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Cumulative ratio of resource	0.00	0.01	0.02	0.04	0.07	0.14	0.25	0.39	0.58	0.78	1.00



- The Gini Coefficient (GC) associated with a given Lorenz curve L(x) is twice the area between y = L(x) and y = x.
- Why twice the area? This produces GC's between 0 and 1. Some authors refer to the GC as the percentage of inequality in the population.
- Advantages:
  - The 'gold standard.'
  - Visually intuitive
  - Takes into account all the data
- Disadvantages:
  - Not as intuitive as some other measures.
  - Calculation is cumbersome.
  - Obscures absolute values.
  - Not having complete individual data.

How do you calculate the area?

Cumulative ratio of pop.	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Cumulative ratio of resource	0.00	0.01	0.02	0.04	0.07	0.14	0.25	0.39	0.58	0.78	1.00



- If you have an equation for the Lorenz curve, use calculus (if you know how).

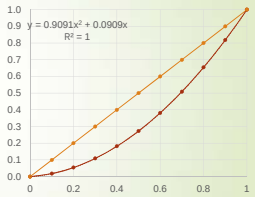
$$GC = 2\int_0^1 (x - L(x))dx$$

$$= 1 - 2\int_0^1 L(x)dx$$

- As a rough estimate, graph the curve on Excel and count grid squares. I get about 11.5 rectangles for a GC of about 0.46
- As an even rougher eye ball estimate, draw a triangle with vertex (1,1) and base along the x-axis starting at 0 with what seems to you to be the same area. The length of the basis is an estimate for the GC.
- Use technology (spreadsheet, online calculator, etc.). Using a trapezoid rule and Excel, I got GC = 0.444

Three examples with (essentially) the same Gini Coefficient

- The bottom 0.30 of the population has none of the resource, and the other 0.70 share the resource equally.
- In a large population, the top individual has 0.30 of the resource and everyone else shares the remaining 0.70 of the resource equally.
- Each decile has k times the resource share of the first decile for k from 2 through 10. This means the bottom decile has 1/55th of the resource.
- The moral of the story: Different Lorenz curves can result in the same GC.



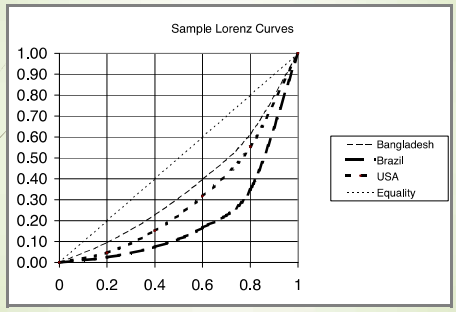
Some national Gini coefficients

Country	UN Gini Coefficient	Rank
Namibia	0.743	1
Sierra Leone	0.629	3
Haiti	0.592	7
South Africa	0.578	10
China	0.469	34
Singapore	0.425	51
United States	0.408	58
India	0.368	76
United Kingdom	0.360	83
Australia	0.352	86
Japan	0.249	110
Denmark	0.247	111

- Data on Gini Coefficients for Nations, U.S. States, and other political or geographic divisions are fairly readily available.
- Often the same sources have data on other social and economic variables.
- Using these data, students in QL (or statistics) courses can consider (possibly) interesting questions. For example . . .

https://www.intmath.com/blog/mathematics/the-gini-coefficient-of-wealth-distribution-4187

4. Income Inequality in Three Countries. The graph below shows the Lorenz curves for Bangladesh, Brazil, and the United States for the year 1989.



(a) Give the percentage of total income earned by the bottom half of the population in each of the three countries.

(b) Give the percentage of total income earned by the *top* 40% of the population in each of the three countries.

(c) Calculate an estimated Gini coefficient for each country, based on the graphs. As noted in the text, the Gini coefficient is twice the area between the equality line and the graph.

(d) Of the three countries for which data is given, Brazil has the highest level of inequality and Bangladesh the least. However, one drawback of the Gini coefficient is that it does not take into account the overall level of income. Per capita incomes around this time for Bangladesh, Brazil, and the U.S. were approximately $150, $2500, and $17,000 respectively. The populations for these countries (in millions) were approximately 109, 148, and 255.

i. Based on the additional data, find the total amount of income earned in each country.

ii. Using the Lorenz curves, find the total amount earned by the bottom 40% of the population in each country. About how much is this per person?

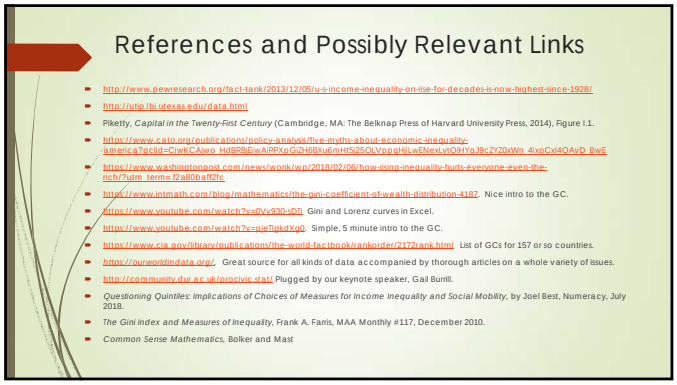
iii. Find the total amount earned by the top 40% of the population in each country. About how much is this per person?

Other Ideas for Using Lorenz Curves and Gini Coefficients

- Consider questions like:
  - Can you create to Lorenz curves with the same GC? Can these intersect? More than once?
  - Are societies which are high in inequality also high in poverty, political turmoil, health problems, crime, etc.?
  - How has inequality changed over time in various countries, states, or cities?
  - Are democratic societies higher or lower in inequality than other societies?
  - What is a permissible level of inequality?
  - How should we talk about inequality? (e.g. "Dream-horders")

Sources of Gini Coefficient Data.

- World Income Inequality database (WIID) from United Nations University. Over 8000 entries covering nearly all countries multiple times. https://www.wider.unu.edu/project/wiid-world-income-inequality-database
- The dreaded Wikipedia. Refers to World Bank and CIA data for GC's, along with some percentile ratios. https://en.wikipedia.org/wiki/List_of_countries_by_income_equality
- OECD Income Distribution Database (IDD). http://www.oecd.org/social/income-distribution-database.htm
- University of Texas Inequality Project. https://utip.lbj.utexas.edu/
- Gini Coefficients by U.S. State. https://www.worldatlas.com/articles/us-states-by-gini-coefficient.html
- Percentile ratios for 100 largest U.S. cities. East Lansing is listed as 4th worst. https://www.brookings.edu/research/city-and-metropolitan-income-inequality-data-reveal-ups-and-downs-through-2016/
- America's Most Unequal Cities by Income. https://www.nerdwallet.com/blog/studies/cities-where-income-is-most-equal-2015/ https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html



References and Possibly Relevant Links

- http://www.pewresearch.org/fact-tank/2013/12/05/u-s-income-inequality-on-rise-for-decades-is-now-highest-since-1928/
- http://utip.lbj.utexas.edu/data.html
- Piketty, *Capital in the Twenty-First Century* (Cambridge, MA: The Belknap Press of Harvard University Press, 2014), Figure I.1.
- https://www.cato.org/publications/policy-analysis/five-myths-about-economic-inequality-america?gclid=CjwKCAiwo_HdBRBiEiwAIPPXpGiZH6BXu6mHt5jz25OLVppqHjLwENexLytQlHYqJ9cZY20xWn_4IxoCxI4QAvD_BwE
- https://www.washingtonpost.com/news/wonk/wp/2018/02/06/how-rising-inequality-hurts-everyone-even-the-rich/?utm_term=.f2a80baff2fc
- https://www.intmath.com/blog/mathematics/the-gini-coefficient-of-wealth-distribution-4187. Nice intro to the GC.
- https://www.youtube.com/watch?v=0Vv930-sDTI Gini and Lorenz curves in Excel.
- https://www.youtube.com/watch?v=pjeTlgkdXg0. Simple, 5 minute intro to the GC.
- https://www.cia.gov/library/publications/the-world-factbook/rankorder/2172rank.html List of GCs for 157 or so countries.
- *https://ourworldindata.org/*. Great source for all kinds of data accompanied by thorough articles on a whole variety of issues.
- http://community.dur.ac.uk/procivic.stat/ Plugged by our keynote speaker, Gail Burrill.
- *Questioning Quintiles: Implications of Choices of Measures for Income Inequality and Social Mobility*, by Joel Best, Numeracy, July 2018.
- *The Gini Index and Measures of Inequality*, Frank A. Farris, MAA Monthly #117, December 2010.
- *Common Sense Mathematics*, Bolker and Mast