The Gini Index: Measuring Inequity

The distribution of income in our society is a concept of ongoing interest to economists, politicians, public policy analysts, and other concerned individuals. The data that economists use to quantify distribution of income is often presented in the form of Table 1, which in this case is constructed using data from 2000. Students can find this data by searching the internet at http://www.census.gov/hhes/www/incineq.html.

Fifth of Households	Percent of income
Lowest fifth	**3.6**
Second fifth	**8.9**
Third fifth	**14.9**
Fourth fifth	**23.0**
Highest fifth	**49.7**

Table 1: Percent distribution of aggregate income for 2000

Table 1 gives the percent of the total income of the United States earned by each fifth of the population, ordered by income. The procedure for determining the numbers in Table 1 can be thought of as follows.

> Each family and each unattached individual counts as one household with one income level. Suppose all families and unattached individuals are lined up according to their earnings for the year, from least income to greatest income. Starting at the least income level, we count off one-fifth of the total number of households. All of the households in this group fall into the lowest fifth in income level.

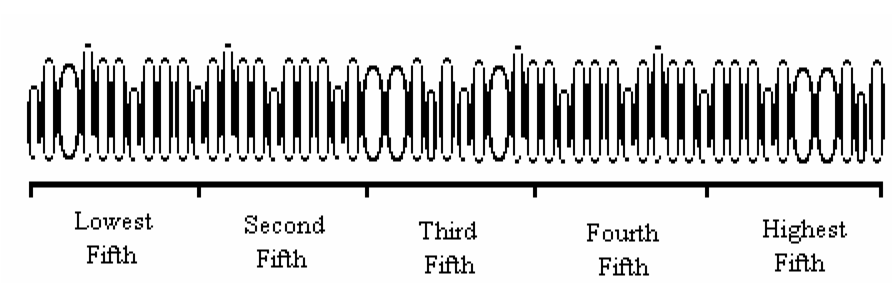


Figure 1: Lineup of population from least income to greatest income

We continue counting households in order of increasing income and divide the number of households into the second, third, fourth, and highest fifth. The total income of each fifth, expressed as a percentage of the total income of all households, gives the percentage distribution for each fifth.

The data in Table 1 mean that in 1994 the lowest fifth of all households earned 4.2% of all the income earned for that year. The second fifth earned 10.0% of all income, while the highest fifth of all households earned 46.9% of all income.

Clearly, the distribution of income is not equitable. How can you measure just how unequitable it is? Given data about income distribution in two different years, we would like a numerical way to compare the extent to which one income distribution is more or less inequitable than another.

Measuring Inequity

Economists have developed a very intuitive measure of this inequality of the distribution of incomes by considering the data that represents the cumulative income. The cumulative data can be obtained by adding successive entries in the right-hand column of Table 1, yielding the data shown in Table 2. Of course the lowest five-fifths of the population, which is all households, earns 100% of the income, although the numbers in Table 1 do not add to exactly 100% because of rounding.

Fifths of Households	**Percent of income**
Lowest one-fifth	**3.6**
Lowest two-fifths	**12.5**
Lowest three-fifths	**27.4**
Lowest four-fifths	**50.4**
Lowest five-fifths	**100.0**

Table 2: Cumulative percent distribution of aggregate income for 2000

A graph of the data in Table 2 is obtained by plotting the cumulative proportional distribution of aggregate income versus the proportion of the population, as shown in the graph given in Figure 2. The percentages have been converted to decimal numbers, that is, the lowest two-fifths earning 12.5% of the aggregate income is represented by the point $(0.4, 0.125)$. The points (0,0) and (1,1) have been included because 0% of the households earn 0% of the income and 100% of the households earn 100% of the income.

We can develop a measure of inequity by comparing this data set to that the corresponding data for a perfectly equitable economy? If everything was equitable, then each fifth of the population would earn one-fifth of the income. The cumulative graph would be that of $y = x$, as shown in Figure 3.

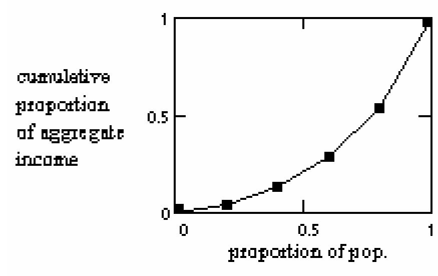


Figure 2: Graph of cumulative proportional distribution data

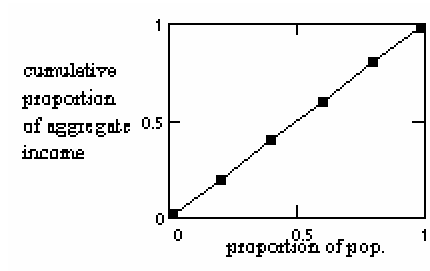


Figure 3: Graph of perfectly equitable cumulative distribution data

How "far" from the graph of perfect equity shown in Figure 3 is the graph of the income distribution for the US in 2000 shown in Figure 2? Students will come up with several nice ways to measure the difference in the two data sets, and therefore the inequality of the income distribution.

The Gini Index

One measure employed by the economists is the ratio of the areas *A* and *B* shown in Figure 4.

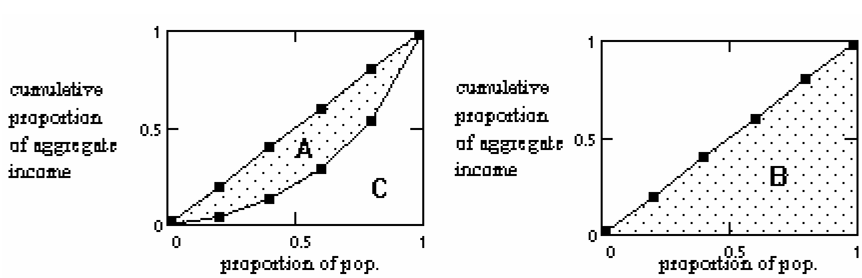


Figure 4: Areas to be computed

This ratio can have a value anywhere from 0, representing perfect equity, to 1, representing perfect inequity. The larger the ratio, the more inequitable the distribution of income. The area under the totally equitable distribution, *B*, is always one-half. To find the area of the shaded region *A*, we need to subtract the area labeled *C* from one-half. How can we find the area under the distribution curve for 2000? How you go about this computation depends upon the level of mathematical experience of the students.

Teachers of calculus will realize that the problem we have is a good example of the area between two curves.

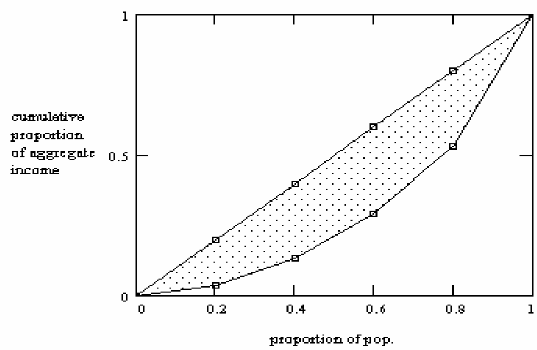


Figure 6: Area between two curves

A curve that models data of the type (proportion of households, cumulative proportion of aggregate income) is called a ***Lorenz Curve.*** Such Lorenz curves always lie somewhere between the two curves $y = x$ and $y = 0$ since income distribution must fall somewhere between perfect equity and perfect inequity.

There are many ways of finding the equation for the Lorenz curve based on the data from Table 2. A reasonable model for this data is a power function of the form $y = x^n$, with $n > 1$. The graph of this function contains the required points (0,0) and (1,1). We choose not to use a power least squares procedure to fit a power function to the data because a Lorenz curve must contain the point (1,1) and a power least squares curve does not necessarily contain (1,1). Instead, we will use the fact that a log-log re-expression linearizes data that is modeled by a power function. Since $y = x^n$, we take the logarithm of both sides of the equation to obtain $\ln y = \ln x^n$,which simplifies to $\ln y = n \ln x$.

Solving for n gives $n = \dfrac{\ln y}{\ln x}$. We can obtain a reasonable estimate for n by forming the ratio $\dfrac{\ln y_i}{\ln x_i}$ for each point (x_i, y_i) in the data set and then averaging these ratios.

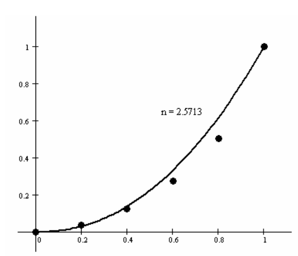
We could also use this problem to have students devise a least-squares estimate. Consider $Y = nX$. We want to minimize $S = \sum_{i=1}^{4} \left(Y_i - nX_i \right)^2$. This is a regression problem in which there is only one variable, so we don't have to use partial derivatives without having defined them.

$$\frac{dS}{dn} = \sum_{i=1}^{4} 2\left(Y_i - nX_i\right)\cdot\left(-X_i\right). \text{ If } \frac{dS}{dn} = 0 \text{ then } \sum_{i=1}^{4} X_i Y_i = n \sum_{i=1}^{4} X_i^2 \text{ and } n = \frac{\sum_{i=1}^{4} X_i Y_i}{\sum_{i=1}^{4} X_i^2}$$

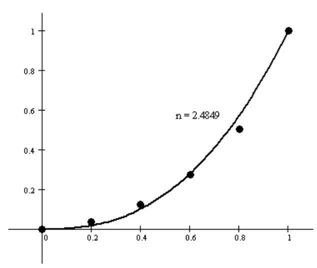
Since $X_i = \ln\left(x_i\right)$ and $Y_i = \ln\left(Y_i\right)$, we have $n = \dfrac{\sum_{i=1}^{4} \ln\left(x_i\right)\cdot\ln\left(y_i\right)}{\sum_{i=1}^{4} \left[\ln\left(x_i\right)\right]^2} \approx \dfrac{\sum_{i=1}^{4} \ln\left(x_i\right)\cdot\ln\left(y_i\right)}{3.7406}$.

Or, we can fit a true least squares model.

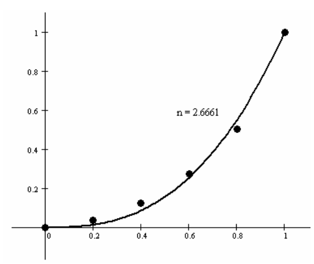
$S = \sum \left(y_i - x_i^{\,n} \right)^2$, so $\dfrac{dS}{dn} = \sum 2\left(y_i - x_i^{\,n}\right)\left(-x_i^{\,n}\right)\ln\left(x_i\right)$ and $0 = \sum \left(y_i - x_i^{\,n}\right)\left(-x_i^{\,n}\right)\ln\left(x_i\right)$, which requires Newton's Method or similar numerical technique.



Least Square on $\ln(y) = n \cdot \ln(x)$



Average $n = \dfrac{\ln(y)}{\ln(x)}$



Least Squares on $y = x^n$

Your students may make another choice for the exponent. As long as they are consistent in their procedure, important comparisons can be made.

In general,

Gini index

$$= \frac{\text{area bounded by Lorenz curve and } y = x}{\text{area of triangle for perfect equity}}.$$

$$= 2\int_0^1 x - x^n dx = 1 - \frac{2}{n+1}$$

For the year 2000, for which the Lorenz curve has power $n = 2.1573$, we have $2\int_0^1 x - x^{2.1573} dx = 1 - \frac{2}{3.1573} \approx 0.3665$.

The value will different according to the method used to find the Lorenz curve. This difference is not important. What we are interested in is comparing the values from year to year using one of the methods. To compare values generated by different methods will give no useful information. The Gini indices over time illustrate the changes in the division of income.

The data in the following table is the income distribution by fifths in the United States for various years since 1947.

Percent Distribution of Aggregate Income

Year	Lowest Fifth	Second Fifth	Third Fifth	Fourth Fifth	Highest Fifth
1947	5.0	11.9	17.0	23.1	43.0
1950	4.5	12.0	17.4	23.4	42.7
1955	4.8	12.3	17.8	23.7	41.3
1960	4.8	12.3	17.8	24.0	41.3
1965	5.2	12.2	17.8	23.9	40.9
1970	5.4	12.2	17.6	23.8	40.9
1975	5.4	11.8	17.6	24.1	41.1
1980	5.1	11.6	17.5	24.3	41.6
1985	4.6	10.9	16.9	24.2	43.5
1990	4.6	10.8	16.6	23.8	44.3
1994	4.2	10.0	15.7	23.3	46.9

Source: Weinberg, Daniel H. "*A Brief Look at Postwar U.S. Income Inequality*, *Current Population Reports: Household Economic Series*" US Census Bureau, June 1996

More information can be found with gender and race distributions at
http://www.census.gov/prod/2000pubs/p60-204.pdf and
http://www.census.gov/hhes/income/midclass/middata.html.

Figure 7 shows the Gini index computed using trapezoids, by calculus, and using the Census Bureau calculations which involves more data.

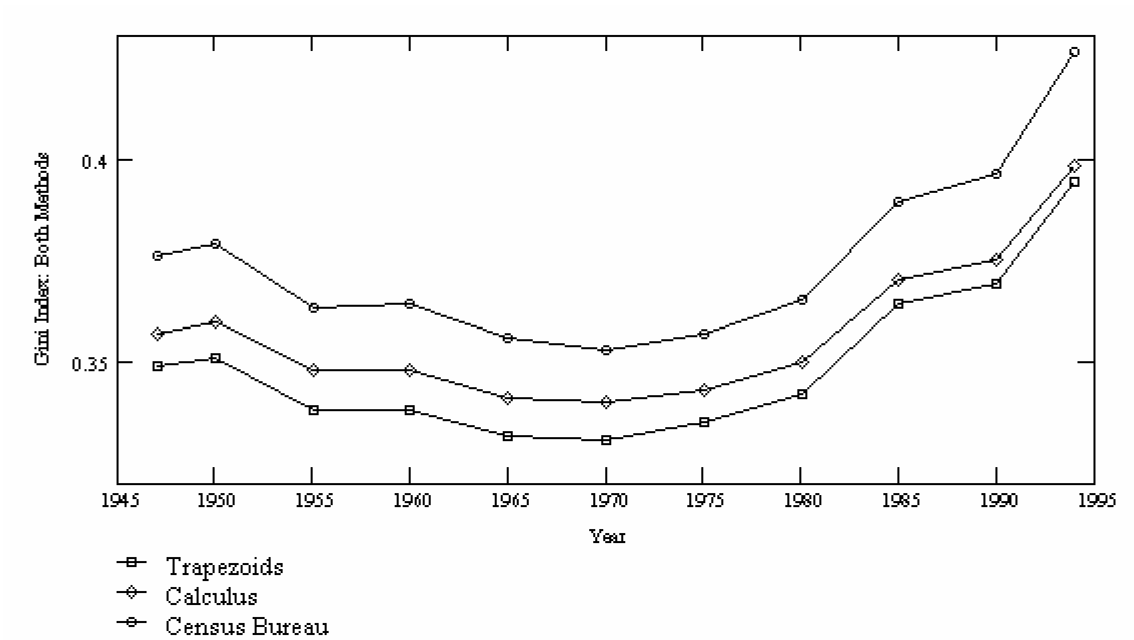


Figure 7: Gini indices over time

Students will notice that while the values of the indices differ using the different methods, the relative values remain quite consistent. If students chose to just compare the areas between $y = x$ and $y = x^n$, or to measure equity rather than inequity, they would always find that the distribution in 1990 was more inequitable than that of 1970.

Student Investigations

One important investigation is the comparison of student measures to the traditional Gini index. Students may prefer their own measure as being more reasonable or simpler. By working out their own method first, the Gini index used by the economists should make good sense. Students can also use the data given to investigate the relative values of the Gini indices for years when the president is a Democrat and for years when the president is a Republican. Our students have enjoyed investigating the historical events leading to the most drastic changes in the Gini index. Another variation that has been used in our government classes has been a comparison of Gini indices for different countries around the world. A Google search on Gini Index turns up all kinds of interesting information and leads to many different student investigations. The comparison of indices for countries with socialistic-style governments will be quite different from those with democratic-style governments. Gini indices for land ownership is often of interest as well. The problem of inequity can easily lead to student projects involving both the mathematics and social studies departments.