Article from George Selitto

– TIME 2004 Presentation – Montreal, July 2004

In 1989, 463,664 students sat for Advanced Placement examinations. This number rose to 1,016,657 in 1998 (The College Board, 1998). During that same period of time the number of students taking the Advanced Placement Calculus Examination has increased every year, from 73,647 in 1989 to 144,759 in 1998 (The College Board, 1999). This expanding interest suggests that many teachers will have in their freshman and sophomore mathematics classes students who will go on to take this examination. The National Council of Teachers of Mathematics in its *Standards* (1989) suggested all students, including those mathematically gifted, would benefit from a curriculum that used technology appropriately while developing in each student an appreciation and facility for estimation. Thus, mathematics teachers are challenged to find ways to apply the *Standards* in ways that will help stimulate student interest in mathematics and also prepare students for the demands of the Advanced Placement Calculus class.

Introducing students to the study of interesting "real-world" applications of mathematics is a tool that teachers can use to accomplish these goals. Asking students to find the area of a region bounded by two curves is one such "real-world" question that can require students to utilize technology in order to answer a "real-world" problem via estimation methods. Specifically, students will be shown how to measure how equitably income is distributed in a population. This will be accomplished by introducing students to two concepts from the field of economics, the Lorentz Curve and the Gini Coefficient. Some time should be spent discussing these concepts with students prior to any actual presentation of data. A perfectly equitable society can be defined as one in which each person in the society earns the same amount of money. Said another way p% of the total income will be earned by p% of the population. In reality, some members of the society will earn more, and at times, far more than their share, while others will earn less.

A simple example is used to illustrate the construction of a Lorenz curve for an imaginary society. Students are presented with a population consisting of 10 people whose incomes are given in Table I below. Income data are listed in increasing numerical order.

Table I

Person ID #	1	2	3	4	5	6	7	8	9	10
Income (in thousands)	6	10	15	20	30	40	52	54	60	80

Using the data in Table I, students generate a set of ordered pairs, (X_n,Y_n), where

X_n = cumulative fraction of the population represented by persons 1 through n, and Y_n = fraction of the total income earned by persons 1 through n. In the example above, first compute the total income for the society, that is, $367 (in thousands). Person #1, who represents .1= 10% of the population, earns 6/367 or .016 = 1.6% of the total income of the society. The ordered pair (X_1, Y_1) = (.1 , .016) where data are left in decimal form. Persons 1 and 2 together represent .2 = 20% of the total population, and they earn 16/367 or .044 = 4.4% of the total income of the society. The ordered pair (X_2, Y_2) = (.2, .044). This process is completed for the remainder of the society, generating a set of ten ordered pairs. The data for the example given above are presented in Table II below.

Table II

Ordered Pair			Ordered Pair	
(X_1,Y_1)	(.1, .016)		(X_6,Y_6)	(.6, .330)
(X_2,Y_2)	(.2 , .044)		(X_7,Y_7)	(.7, .471)
(X_3,Y_3)	(.3 , .085)		(X_8,Y_8)	(.8, .619)
(X_4,Y_4)	(.4 , .139)		(X_9,Y_9)	(.9, .782)
(X_5,Y_5)	(.5, .221)		(X_{10},Y_{10})	(1.0, 1.00)

It should be pointed out that if this society was perfectly equitable, then $X_n = Y_n$ for each data point. That is, in a perfectly equitable society, 30% of the population simply would share 30% of the income of that society.

After all points are plotted, they are connected using a smooth, continuous curve. Economists call this curve the Lorenz Curve. The exact curve through the points may be very complex and difficult to compute. Thus, this curve is usually estimated using regression techniques. Many different technologies can be used to approximate the Lorentz Curve. Often students will be presented with two or three choices that look good graphically. On what basis can they make their choice? One easily understood method to make this determination is to compare the sum of the absolute differences between the observed fraction of total income and the fraction of total income predicted by the approximation of the Lorentz Curve, $|O_j - P_j|$. Students easily understand, and thus accept, the fact that the curve with the smallest $\Sigma|O_j - P_j|$ is the curve with the best fit. (The reader will note that this lesson also prepares students for idea of least squares.)

At this point, it is important to explain and illustrate what the Lorenz curve would look like for a society in which income were distributed perfectly equitably. Students need to be able to clearly articulate why the Lorenz curve would be the line y = x if income were equitably distributed. The graph of the line y = x is then plotted on top of the Lorenz curve for the data set in the example given. A discrepancy is clearly evident between the line y = x and the regression curve approximating the Lorenz curve. It is just this area between the two curves that addresses the question, “What is the measure of income inequity in a society?” The greater the area between the Lorenz curve and the graph of the line y = x, the more inequitable is the distribution of income for the society.

A natural next step for student exploration is to attempt to quantify the inequitable distribution of income for the society. In order to do this it is necessary to measure the area between the Lorenz curve formed above and the graph of the line y = x. Economist call this quantified measure the Gini coefficient. The Gini coefficient is computed mathematically by dividing the area between the Lorenz curve and the line y = x by the area between the graph of the line y = x and the x-axis over the interval [0,1]. It is obvious that the area between the graph of the line y = x and the x-axis over the interval [0,1] is ½ or .5 since a right isosceles triangle is formed with side = 1. Therefore the Gini coefficient is equal to (Area between y = x and L(x)) divided by ½, or 2 X (Area between y = x and L(x)).

Students should be challenged to explain why a Gini coefficient of 0 would correspond to a perfectly equitable distribution of income for the society, while a Gini coefficient of 1 would correspond to a perfectly inequitable society. If the Gini coefficient is 0 then the area between the line y = x and the Lorenz curve must be 0, or close to it. Students should be asked to draw the curves that would result in a Gini coefficient of 1. They can also be encouraged to do additional research on the concept of the Gini coefficient, its origins, uses, and interpretations in economics.

The Algorithm

The approximation of the Lorenz curve and the Gini Coefficient of the imaginary society from Table I will be calculated using the TI-83.

Step I: Entry of Original Data

1. Enter the Person ID # 's in List 1, $\mathbf{L_1}$ and the corresponding incomes into list 2, $\mathbf{L_2}$.
2. Store the number of people in the sample in **N**, by using the following commands: Largest ID # $\rightarrow$ **N** (In this example **N** = 10.)

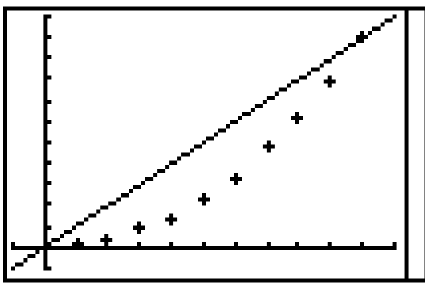
Step II: Calculation of Cumulative Fraction Information

1. Calculate and store the cumulative fraction of population information by using the following command. $\mathbf{L_1/N \rightarrow L_3}$, where N is the number of people in the population.
2. Calculate and store the fraction of the total earned by each individual in the population by using the following command. $\mathbf{L_2/sum(L_2) \rightarrow L_4}$, (The sum(list) command can be found in the MATH submenu of the LIST menu of the TI-83.)
3. Calculate and store the cumulative fraction of income earned by successive groups by using the following commands. $\mathbf{cumSum(L_2)/sum(L_2) \rightarrow L_5}$, (The cumSUM(list) command can be found in the OPS submenu of the LIST menu of the TI-83.)

Step III: A Peek at the Graphs

1. Go to the STAT PLOT menu of the TI-83 and set up the following plot.
   Type: scatter
   Xlist: L_3
   Ylist: L_5.
2. Next set the variable Y_1 equal to X. On the TI-83 this is done using the Y= graphing key found in the top row of keys on the calculator keyboard.
3. Plot the line y=x along with the scatter plot in a window with both the x and y values ranging from -.1 to 1.1. The TI-83 output screen can be seen in Figure I below.

Figure I

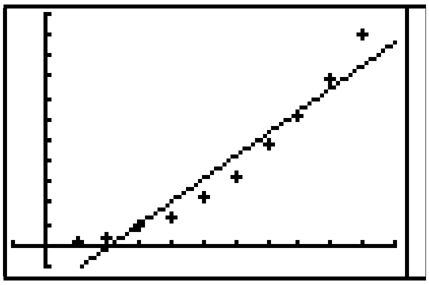


Step IV: Approximating the Lorenz Curve via Regression

Next various curves will be used to approximate the behavior of the actual data. These curves, and the quality of their fit, will be found using several of the built-in procedures of the TI-83.

1. The line which best fits the data is found by using the LinReg(ax+b) procedure found in the CALC sub-menu of the STAT menu of the TI-83. The following commands will find the find the desired line, **LinReg(ax+b) L_3,L_5,Y_2**. (The Y_2 can be found in the Y-VARS sub-menu of the VARS menu of the TI-83.) The resulting line, $y = 1.079x - .223$, is stored as Y_2. The TI-83 output screen can be seen in Figure II below with both x and y ranging from -.1 to 1.1.

Figure II



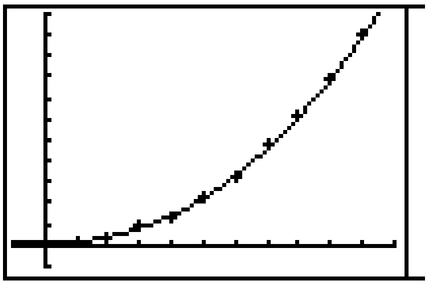
2. The $\Sigma|O_j - P_j|$ is found using the following command **sum(abs(ʟRESID)).** (The sum command is found in the MATH sub-menu of the List menu. The abs command is found in the NUM submenu of the Math menu. The list of residuals, ʟRESID, is found in the NAMES sub-menu of the LIST menu.) The resulting sum of absolute residuals for the linear fit is .719.

3. Repeat steps 1 and 2 for various regression models. These should include quadratic, cubic and quartic along with any other model that is appropriate for the level of the class. The model with the smallest $\Sigma|O_j - P_j|$ will be considered the model with best fit.

Model	$\Sigma\|O_j - P_j\|$
Linear	.719
Quadratic	.054
Cubic	.050
Quartic	.048

Thus, in this example the Quartic equation $-.119x^4 + .392x^3 + .735x^2 - .026x + .013$ is considered the curve of best fit . The TI-83 output screen can be seen in Figure III below with both x and y ranging from -.1 to 1.1.

Figure III



STEP V: Calculating the Gini Coefficient

The Gini Coefficient can be approximated using the following program. This program divides the interval [0, 1] into 10 subintervals and finds the area of the rectangle formed by each subinterval. The fact that this program yields an estimate of the true area between the line y = x and the chosen approximation of the Lorenz Curve should be discussed with the class.

```
:0→S
:For(I, 0, .9, .1)
:Y₁(I) - Y₂(I)
:S + .1(Y₁(I) - Y₂(I))→S
:End
:Disp "SUM:", S
:Disp "GINI COEFF:", 2*S
```

For the example under discussion this program yields a Gini Coefficient of .356. This value suggests a society in which income is moderately well distributed.

Discussion

The introduction of the economics concepts illustrated above introduces students to many ideas they will see in more advanced mathematics classes, shows students that estimation methods are more than guesses and provides a wealth of potential topics for discussion and research. Students who complete and understand the above exercise have been introduced to important ideas from statistics and calculus. They have seen how rectangles can be used to estimate a non-rectangular area. Follow-up exercises, if assigned, may help them understand the effect of dividing up the interval [0, 1] into more than ten sub-intervals or may help them understand the effect of picking something other than the left endpoint of each interval to calculate rectangle heights. Finally, students can be asked to investigate how these economic concepts are used in society today. They can also be encouraged to do some additional research on the history and development of the Lorenz curve and Gini coefficient concepts and share what they find with their classmates at a later date.