Transforming Relationships

About the Lesson

Students will assess the strength of a linear relationship using a residual plot. As a result, students will:

- Calculate the correlation coefficient and coefficient of determination to assess the data set.
- Learn to transform one or two variables in the relationship to create a linear relationship.

Vocabulary

- regression
- residual
- transformation

Teacher Preparation and Notes

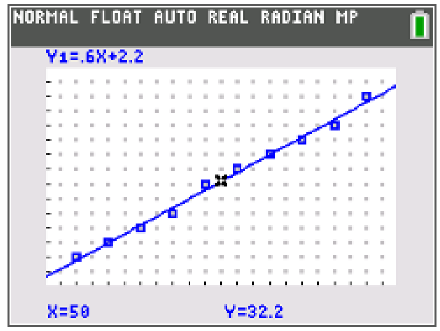
- This activity is designed as a class activity. The teacher document has questions that should be asked to lead the discussion. More background information may be needed for some classes, depending on their knowledge of inverse functions and logarithms.
- The extension can be used or omitted, depending on the ability of the class and the time available.

Activity Materials

- Compatible TI Technologies:

  TI-84 Plus*
  TI-84 Plus Silver Edition*
  TI-84 Plus C Silver Edition
  TI-84 Plus CE

** with the latest operating system (2.55MP) featuring MathPrint ™ functionality.*



Tech Tips:

- This activity includes screen captures taken from the TI-84 Plus CE. It is also appropriate for use with the rest of the TI-84 Plus family. Slight variations to these directions may be required if using other calculator models.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials
- Any required calculator files can be distributed to students via handheld-to-handheld transfer.

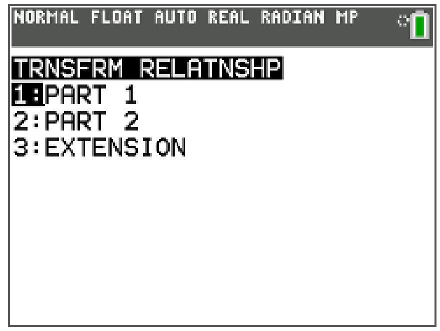
Lesson Files:

- Transforming_Relationships_Student.pdf
- Transforming_Relationships_Student.doc
- DATA.8xp

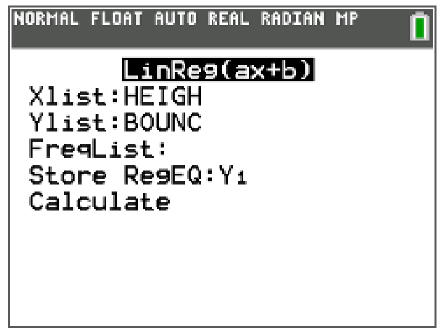
Tech Tip: Before beginning the activity, the program DATA.8xp needs to be transferred to the students' calculators via handheld-to-handheld transfer or transferred from the computer to the calculator via TI-Connect™ CE Software.

Problem 1 – Analyzing Residual Plots

PART 1 in the program **DATA** enters four data sets into the Stat Editor. Students are asked to create a linear regression model and assess the fit graphically and numerically. They are to use Plot1 to create all of the scatter plots. This will ensure that they do not graph two scatter plots at once.



When students choose **LinReg(ax+b)** from the CALC menu, a screen similar to that shown at the right will be displayed. Students should enter the Xlist, Ylist, a location (e.g., $\mathbf{Y_1}$) to store the regression equation, and then select **Calculate**.



Tech Tip: If your students are using the TI-84 Plus CE have them turn on the GridLine by pressing [2nd] [zoom] to change the [format] settings. If your students are using TI-84 Plus, they could use GridDot.

1. Fill in the chart below.

Answers:

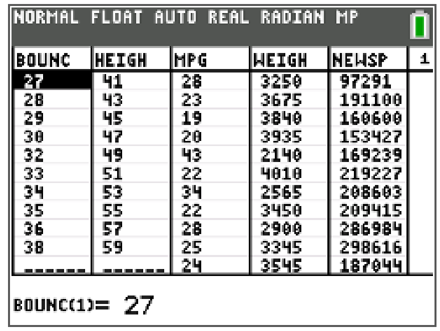
	Independent Variable	Dependent Variable
Bounce and Height	Height	Bounce
MPG and Weight	Weight	MPG
Tons of paper and Year	Year	Tons of paper
Population and Year	Year	Population

2. What is your initial impression of how the regression line fits the data?

Sample Answer: The regression line seems to fit the data. It goes through a couple of points.

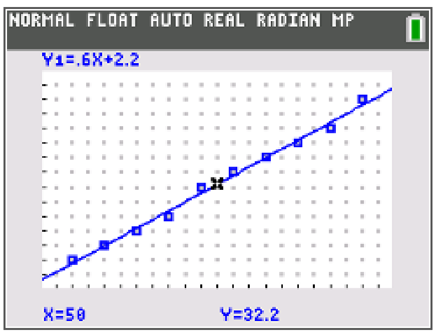
The first two data sets are obviously linear. For the third, linear is not a good fit but the residuals do not show a curved pattern. While the linear model is not good, there is still no other model that works better. Discuss this point with students. In addition, the *r*-value in graph four is good, but the residual plot shows a curved pattern.

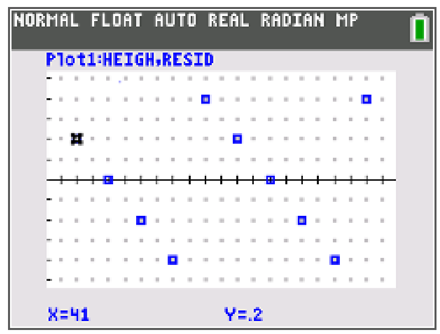
Note: Pressing [2nd] [entry] multiple times will recall the last few entries on the Home screen.



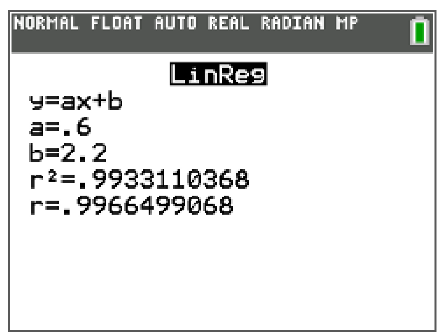
BOUNC	HEIGH	MPG	WEIGH	NEWSP
27	41	28	3250	97291
28	43	23	3675	191100
29	45	19	3840	160600
30	47	20	3935	153427
32	49	43	2140	169239
33	51	22	4010	219227
34	53	34	2565	208603
35	55	22	3450	209415
36	57	28	2900	286984
38	59	25	3345	298616
------	------	24	3545	187044

Tech Tip: Students can quickly turn on the stat diagnostics by pressing [mode] and use the down arrow key to move to STAT DIAGNOSTICS. Press [enter] when ON is highlighted.





Point to emphasize: A low *r*-value means that there is a poor fit in a linear model. Together, the residual plot and the *r*-value help determine if a non-linear model is more appropriate.



More discussion can take place as to what each regression line specifically means. For example, the regression line with bounce vs. height tells us that they are related linearly. Specifically, the regression equation tells us that for each additional unit of height, the bounce increases by 0.6 units.

3. Assess the quality of the fit. Explain your reasoning.

Answer: The linear regression line fits the data because there is no obvious curved pattern in the graph of the residuals. Some points are above the *x*-axis and others are below.

4. What are the values of *r* and r^2? What do these values tell you?

Answers: $r = 0.99665$ and $r^2 = 0.993311$; These values tell you that the data can be modeled by a linear relationship since they are close to 1.

5. How well did a linear model fit the **BOUNCE VS. HEIGHT** graph? Explain your reasoning.

Sample Answer: The linear model fit well because the residual plot, *r*, and r^2 all support this.

6. Interpret the regression equation. What does it specifically tell us about the relationship between drop height and bounce?

Answer: For every additional unit increase in height, the bounce increases by 0.6 units.

7. Now, graph and analyze the other three data sets and fill in the chart below.

Answers:

	Graphically	Numerically
Xlist: WEIGH **Ylist: MPG**	Even distribution above and below. No obvious pattern.	$r = -0.959483$ $r^2 = 0.920607$
Xlist: YEAR **Ylist: NEWSP**	Even distribution above and below. No obvious pattern.	$r = 0.576045$ $r^2 = 0.331828$
Xlist: POPYR **Ylist: POP**	Obvious curved pattern.	$r = 0.95641$ $r^2 = 0.914742$

8. Which of these data sets appears to have a relationship that is non-linear?

Answer: Population vs. Population Year

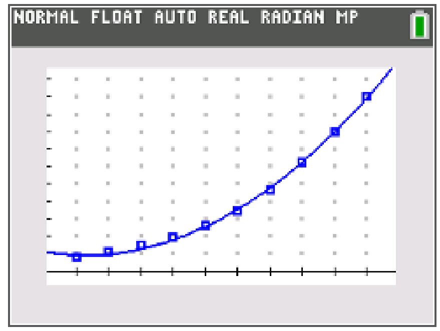
Problem 2 – Transforming Data

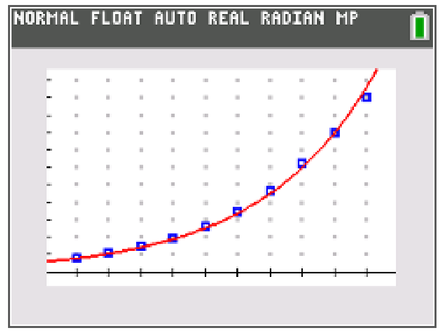
Students will now focus on the Population and Year data set because a linear model did not fit the data.

Allow students the opportunity to play with the many models available and explain which model they feel best represents the data. Some may think that a quadratic model is best, while others might think that an exponential model is more appropriate. Ultimately, they should see that a quadratic model may not be ideal because the residual plot is not random.

Discussion can focus on population growth and exponential modeling.

The screen shots at the right show the quadratic model (**QuadReg**) on the top and the exponential model on bottom (**ExpReg**).





9. What type of graph would model the data set you chose in Question 8? Why?

<u>Answer:</u> Exponential because the data increases slowly at first but then increases rapidly.

10. Try other regressions from the list in the [stat] CALC menu. Which do you feel is the best for this data set? Why?

<u>Sample Answer:</u> Exponential because the graph goes through all but one data point.

After the exponential model is agreed upon, guide students in a discussion on how to transform the data and why a linear model will fit the transformed data. You may need to guide the discussion towards a logarithmic function (which is the inverse of an exponential function).

Students are to calculate the logarithm of the population data. Then they are to graph **LGPOP vs. POPYR** and carry out the assessment of the linear model.

NORMAL FLOAT AUTO REAL RADIAN MP

POPYR	POP	LGPOP	------	------
1790	3.9	.59106		
1800	5.3	.72428		
1810	7.2	.85733		
1820	9.6	.98227		
1830	12.9	1.1106		
1840	17.1	1.233		
1850	23.2	1.3655		
1860	31.4	1.4969		
1870	39.8	1.5999		
1880	50.2	1.7007		
------	------	------		

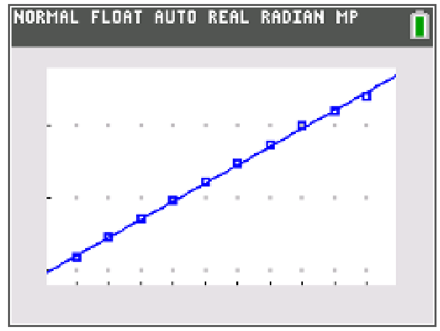
LGPOP(1)=.5910646070265

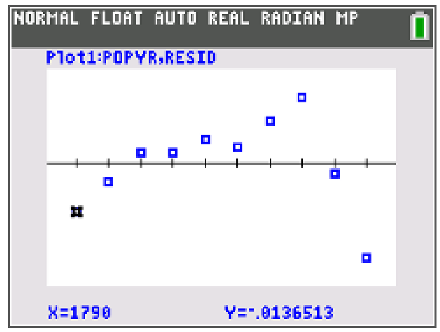
Teacher Tip: Good questions to ask:

- *How does the inverse of a function relate to the original function?*
- *What is the inverse of an exponential function?*
- *Will the base of the logarithm make a difference?*

More discussion can focus on transformations and power functions. Many textbooks contain a more thorough discussion and more background information.

If the linear model fits the transformed graph well, then that means the exponential regression model chosen for the original data fits well.





11. How is this graph different from the original scatter plot **POP vs. POPYR**? What shape is the new graph?

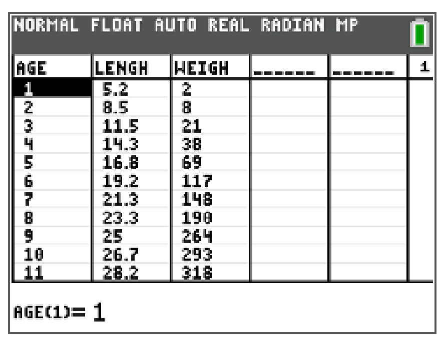
<u>Answer:</u> It appears to be linear.

12. Find the linear regression model and perform the two tests to determine if the data now follows a linear model.

<u>Answer:</u> Graphical: While it looks like there may be a pattern to the residual plot, the values of the residuals are very close to zero. Numerical: r = 0.999402; r^2 = 0.998805

Extension – An Additional Transformation

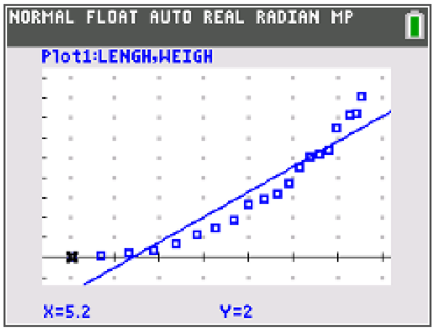
The data given will be best modeled by a power function. This means that both variables will need to be transformed using a logarithmic transformation to create a data set for which students can fit a linear model. Students should be aware that the **WEIGH** list is different from the **WEIGH** list in Problem 1.

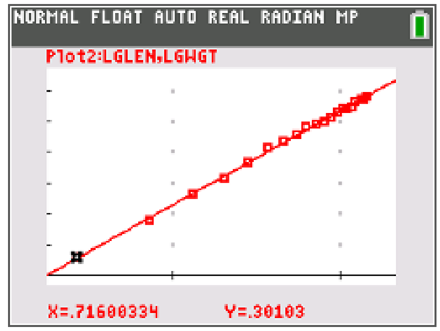


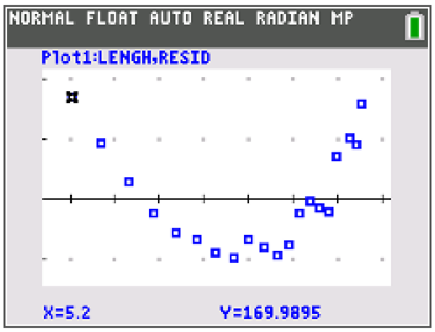
NORMAL FLOAT AUTO REAL RADIAN MP

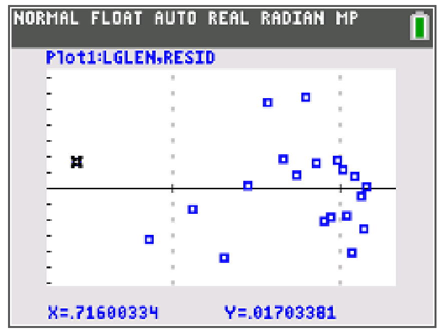
AGE	LENGH	WEIGH	------	------
1	5.2	2		
2	8.5	8		
3	11.5	21		
4	14.3	38		
5	16.8	69		
6	19.2	117		
7	21.3	148		
8	23.3	190		
9	25	264		
10	26.7	293		
11	28.2	318		

AGE(1)= 1









13. Graph **WEIGH vs. LENGH**, add the linear regression line, and graph the residuals. Assess the fit.

Answer: The residuals show a very definite curved pattern; r = 0.946104 and r^2 = 0.895112

14. Test other models. Is there one that works the best?

Answer: Power function

15. Transform the dependent variable, graph the data, and assess the fit.

Answers: The residuals for a linear regression show a very definite curved pattern. This time the curve is upside-down; r = 0.963202 and r^2 = 0.927759

16. Transform the independent variable, graph the two transformed lists against each other, and assess the fit.

Answers: The residuals show no pattern and have very small magnitudes; r = 0.999261 and r^2 = 0.998523

17. What does this tell you about a correct model for these data?

Answer: A power function is a good model for the data.