Logarithmic Transformations of Data

Math Objectives

- Students will understand how various data transformations can be used to achieve a linear relationship and, consequently, how to use the methods of linear regression to obtain a line of best fit.
- Students will use various calculator regression functions in order to model the relationship between two variables.
- Students will convert a linear regression equation back to a non-linear equation to model the relationship between two variables.
- Students will use appropriate tools strategically (CCSS Mathematical Practice).
- Students will look for and make use of structure (CCSS Mathematical Practice).

Vocabulary

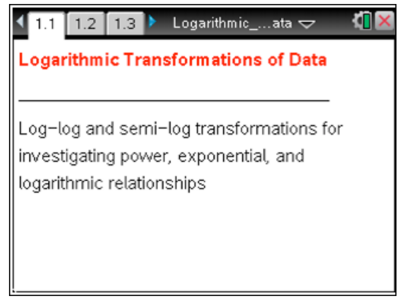
- transformation
- linear regression
- exponential regression
- power regression
- logarithmic regression

About the Lesson

- This lesson involves three real-world data sets in which the relationship between each pair of variables is non-linear.
- As a result, students will:
  - Learn how data transformations can be used to achieve a linear relationship.
  - Learn how to use several calculator regression functions to model the relationship between two variables.
  - Learn how to convert a linear regression equation back to a non-linear equation.

TI-Nspire™ Navigator™ System

- Use Screen Capture to compare the graphs of student regression equations.
- Have students write possible relationships between the original variable pairs on a calculator Notes page.
- Use Screen Capture to compare and discuss student equations.



TI-Nspire™ Technology Skills:

- Download a TI-Nspire document
- Open a document
- Move between pages
- Use various regression functions

Tech Tips:

- Make sure the font size on your TI-Nspire handhelds is set to Medium.
- You can hide/reveal the function entry line by pressing ctrl G.

Lesson Materials:
Student Activity
Logarithmic_Transformations_of_Data_Student.pdf
Logarithmic_Transformations_of_Data_Student.doc
TI-Nspire document
Logarithmic_Transformations_of_Data.tns

Visit www.mathnspired.com for lesson updates and tech tip videos.

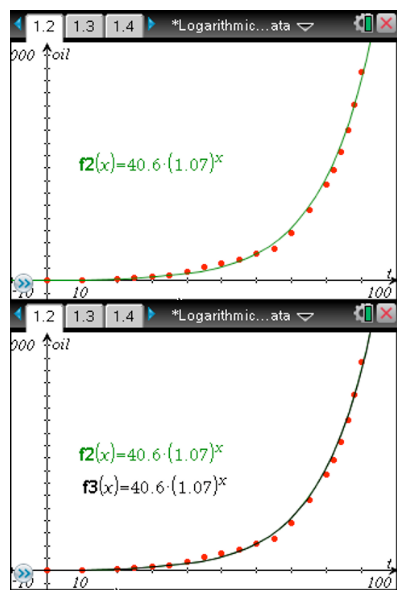
Discussion Points and Possible Answers

Tech Tip: When using one of the calculator regression functions, make sure students enter the appropriate values for X List, Y List, Save RegEqn to, and 1st Result Column. In addition, after a regression equation is saved as a function, the user must trigger the graph of the function on the graph screen on the appropriate Entry Line.

Move to page 1.2.

Teacher Tip: This activity involves several challenging mathematical concepts. For example, students should have knowledge of logarithmic, exponential, and power functions. This activity also requires familiarity with TI-Nspire calculator functions. Instructors might consider working with students to solve Problem 1 and allow students to consider Problems 2 and 3 on their own. In addition, there is an extra Data and Statistics page at the end of each problem. Teachers and students might find it easier to construct scatter plots and find regression equations on these pages. Each Data and Statistics page provides an excellent opportunity to visualize each transformation dynamically.

Move to page 1.2.



1. This problem involves the world production of oil as a function of time[1]. Let t be the time in years since 1880 and the variable oil represents the world production of oil in millions of barrels.
   a. Page 1.2 shows a scatter plot of oil versus t. Describe the relationship between t and **oil**, and write a possible mathematical equation to describe this relationship.

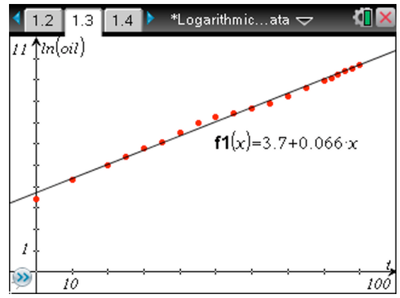
   <u>Sample Answers:</u> The scatter plot suggests there is an exponential relationship between oil and time t. A possible mathematical equation to describe the relationship between **oil** and t is $\text{oil} = a \cdot b^t$, where a and b are constants.

[1] Larson, Pia Veldt. "ST111: Data." May 26, 2003. http://statmaster.sdu.dk/courses/st111/data/index.html. The author cites this source: Joglekar, G., Schuenemeyer, J.H., and LaRiccia, V. "Lack-of-fir testing when replicates are not available." American Statistician, 43, pp. 135-143. 1989.

TI-Nspire Navigator Opportunity: ***Quick Poll (Open Response)***
See Note 1 at the end of this lesson.

Move to page 1.3.

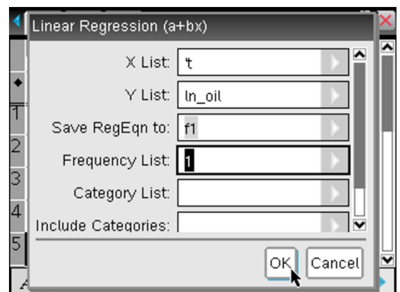
b. This graph is a scatter plot of $\ln(\text{oil})$ versus t. Describe in words the relationship between t and $\ln(\text{oil})$. Describe the effect of this transformation on the relationship between the two variables.

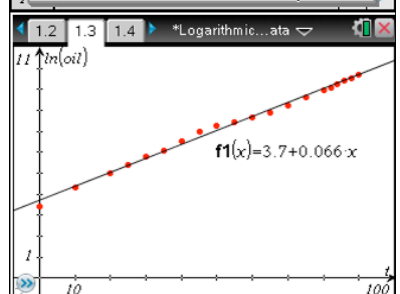


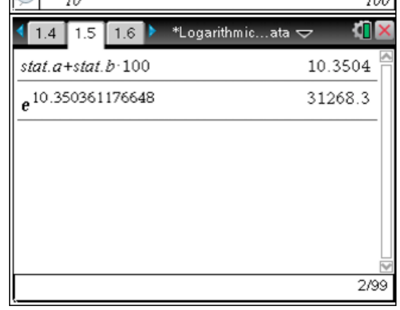
Answer: The scatter plot suggests the relationship between $\ln(\text{oil})$ and t is linear. This transformation appears to achieve a linear relationship between $\ln(\text{oil})$ and time t.

Move to page 1.4.

c. Select **MENU > Statistics > Stat Calculations > Linear Regression (*a*+*bx*)** to find the line of best fit for the ordered pairs $(t, \ln\_\text{oil})$. Store the regression equation in $f1$ and add the graph of $y = f1(x)$ to the scatter plot on Page 1.3. Use this equation to predict the world oil production in the year 1980.







Answer: The regression equation is
$\ln(\text{oil}) = 3.7031 + 0.0665\,t.$

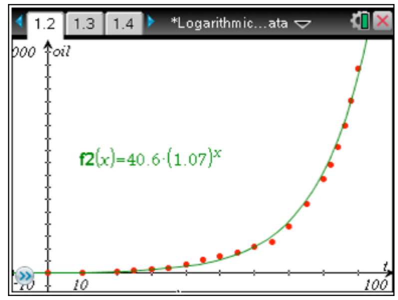
$$1980 - 1880 = 100 = t$$

Insert a calculator page and use the statistics variables generated by the regression function for greater accuracy.

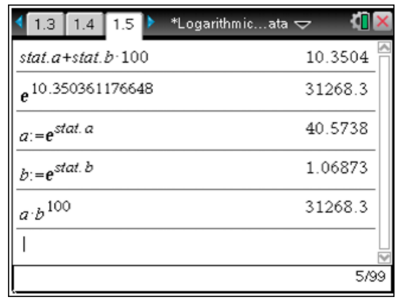
$$\ln(\text{oil}) = 3.7031 + 0.0665\,(100) = 10.3504$$

$$\text{oil} = e^{10.3504} = 31268.3$$

d. Use the regression equation in part (c) to write an equation to describe the relationship between the two original variables, t and **oil**. Add the graph of this function, $y = f2(x)$, to the scatter plot on Page 1.2. Use this equation to predict the world oil production in the year 1980. Does this prediction agree with your answer in part (c)?



Answer: Start with the regression equation, and solve for the variable **oil**. Use a Calculator page with the statistics variables generated by the regression function for greater accuracy.



$$\ln(\text{oil}) = 3.7031 + 0.0665t$$
$$\text{oil} = e^{3.7031+0.0665t}$$
$$= e^{3.7031} \cdot e^{0.0665t}$$
$$= 40.5738 \cdot 1.0687^{t}$$

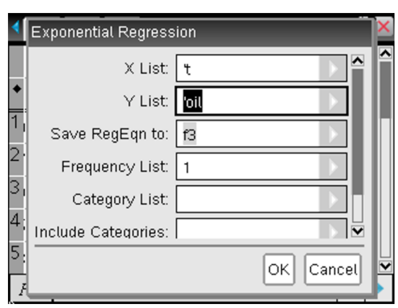
For $t = 100$,

$$\text{oil} = 40.5738 \cdot 1.0687^{100} = 31268.3$$

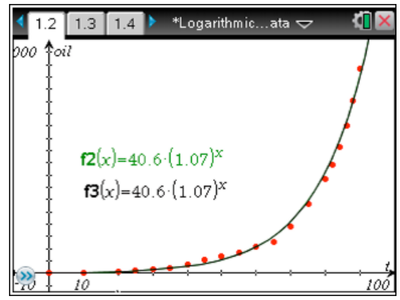
TI-Nspire Navigator Opportunity: *Screen Capture*
See Note 2 at the end of this lesson.

Teacher Tip: These two predictions are the same. Ask students to find the actual world oil production in 1980 and compare this with the predicted value.

e. On page 1.4, select **MENU > Statistics > Stat Calculations > Exponential Regression** to find an equation of best fit for the ordered pairs (t, oil). Store this equation in $f3$ and add the graph of $y = f3(x)$ to the scatter plot on Page 1.2. How do the graphs of $y = f2(x)$ and $y = f3(x)$ compare?

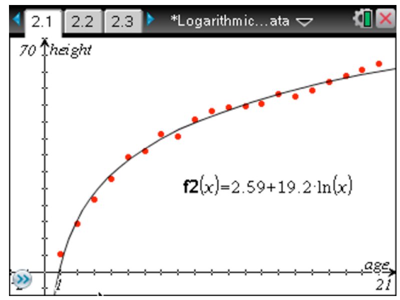


Answer: The two graphs (equations) are exactly the same. Note: Ask students what this suggests about the calculator algorithm for performing exponential regression. (It appears the calculator does the linear transformation in the background first, finds the regression equation, and then converts back to exponential form. This is exactly what we did.)



TI-Nspire Navigator Opportunity: ***Screen Capture***
See Note 3 at the end of this lesson.

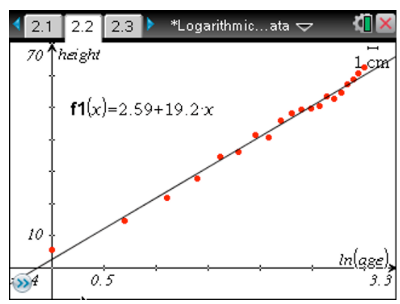
Move to page 2.1.



2. This problem involves the height and age of Douglas Fir trees. Let the variable **height** represent the height of a randomly selected Douglas Fir tree (in feet) and the variable **age** represent the age of that tree (in years).
   a. Page 2.1 shows a scatter plot of **height** versus **age**. Describe in words the relationship between **age** and **height**, and write a possible mathematical equation to describe this relationship.

Answer: The scatter plot suggests there is a logarithmic relationship between height and age. A possible mathematical equation to describe the relationship between height and age is $\text{height} = a + b\ln(\text{age})$ where a and b are constants.

Move to page 2.2.

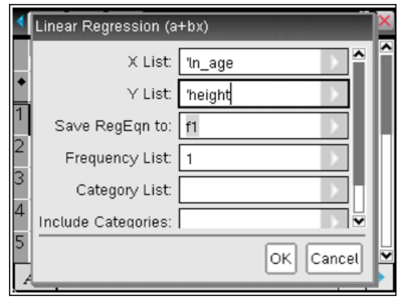


b. This graph is a scatter plot of height versus $\ln(\text{age})$. Describe in words the relationship between height and $\ln(\text{age})$.

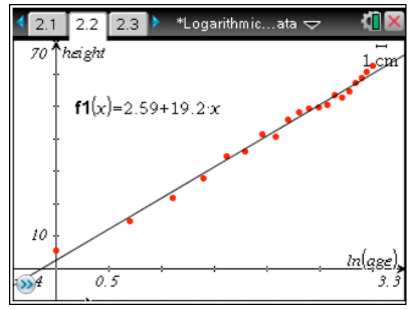
Describe the effect of this transformation on the relationship between the two variables.

Answer: The scatter plot suggests the relationship between height and $\ln(\text{age})$ is linear. This transformation appears to achieve a linear relationship between height and $\ln(\text{age})$.

Move to page 2.3.



c. Select **MENU > Statistics > Stat Calculations > Linear Regression (a+bx)** to find the line of best fit for the ordered pairs $(\ln(\text{age}), \text{height})$. Store the regression equation in $f1$, and add the graph of $y = f1(x)$ to the scatter plot on Page 2.2. Use this equation to predict the height for a 25 year old tree.



<u>Answer:</u> The regression equation is

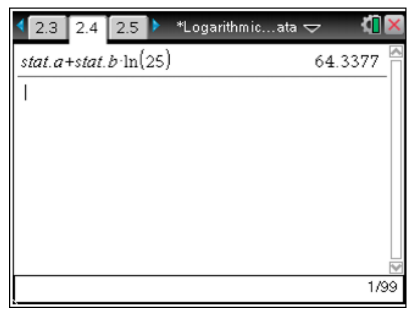
$\text{height} = 2.58656 + 19.1841 \ln(\text{age})$.

$\text{age} = 12$

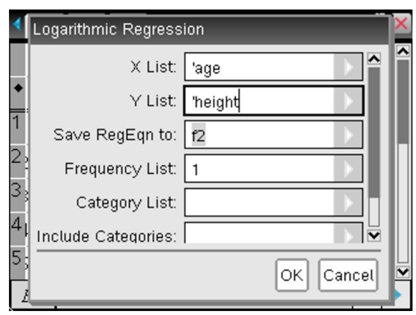
Insert a calculator page, and use the statistics variables generated by the regression function for greater accuracy.

$\text{height} = 2.58656 + 19.1841 \ln(25) = 64.3377$

For a 25 year old Douglas Fir, the predicted height is 64.3377 feet.



d. Select **MENU > Statistics > Stat Calculations > Logarithmic Regression** to find an equation of best fit for the ordered pairs $(\text{age}, \text{height})$. Store this equation in $f2$, and add the graph of $y = f2(x)$ to the scatter plot on Page 2.1. . Use $f2$ to predict the height for a 25 year old tree. How does this prediction compare with your answer in part (c)?



<u>Answer:</u> The logarithmic regression equation is

$\text{height} = 2.58656 + 19.1841 \ln(\text{age})$

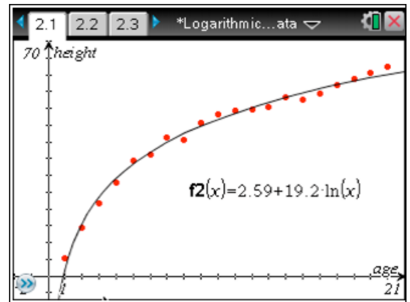
$\text{age} = 12$

Insert a calculator page, and use the statistics variables generated by the regression function for greater accuracy.

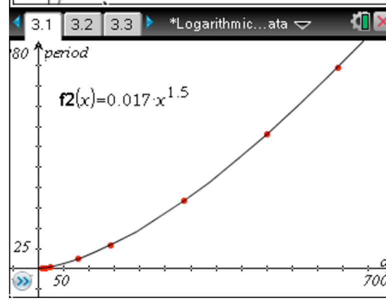
$\text{height} = 2.58656 + 19.1841 \ln(25) = 64.3377$

For a 25 year old Douglas Fir, the predicted height is 64.3377 feet.

This prediction is exactly the same as in part (c).



Move to page 3.1.



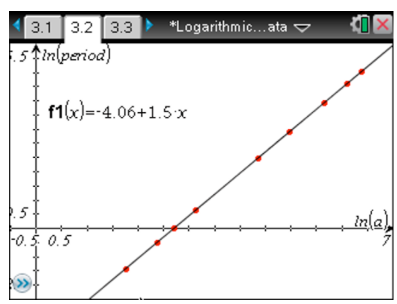
3. This problem involves Kepler's Law of Periods[2], which states that the orbital period of planets revolving about the sun (period) is related to the semi-major axis of its orbit (a).

   a. Page 3.1 shows a scatter plot of period versus a. Describe in words the relationship between **period** and a, and write a possible mathematical equation to describe this relationship.
   Note: You might have to zoom-in and adjust the window to detect the relationship between the variables.

   <u>Answer:</u> The scatter plot suggests there is a power relationship between **period** and a. A possible mathematical equation to describe the relationship between period and a is $\text{period} = c \cdot a^d$ where c and d are constants.

   Note: It might be difficult for students to determine this mathematical relationship from the scatter plot. Consider using the Data and Statistics page at the end of this problem with various variable combinations.

Move to page 3.2.



b. This graph is a scatter plot of $\ln(\text{period})$ versus $\ln(a)$. Describe in words the relationship between $\ln(\text{period})$ and $\ln(a)$. Describe the effect of this transformation on the relationship between the two variables.

   <u>Answer:</u> The scatter plot suggests the relationship between $\ln(\text{period})$ and $\ln(a)$ is linear. This transformation appears to achieve a linear relationship between $\ln(\text{period})$ and $\ln(a)$.

---

[2] Data from http://hyperphysics.phy-astr.gsu.edu/hbase/kepler.html. Additional source cited: Resnick, R., Halliday, D., and Walker, J. Fundamentals of Physics, Fourth Edition. John Wiley & Sons. Table 15-3.

Move to page 3.3.

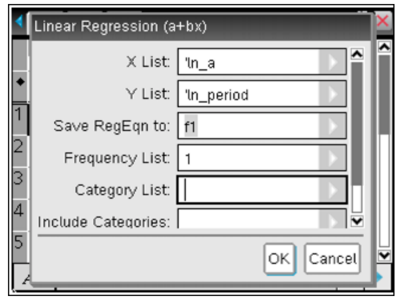
c. Select **MENU > Statistics > Stat Calculations > Linear Regression (a+bx)** to find the line of best fit for the ordered pairs $(\ln(a), \ln(\text{period}))$. Store the regression equation in $f1$, and add the graph of $y = f1(x)$ to the scatter plot on Page 3.2. Use this equation to predict the period for a planet with $a = 100$.

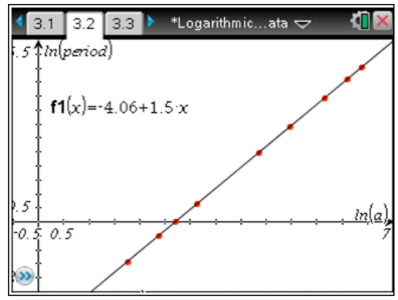
<u>Answer:</u> The regression equation is
$\ln(\text{period}) = -4.05821 + 1.50007\ln(a)$

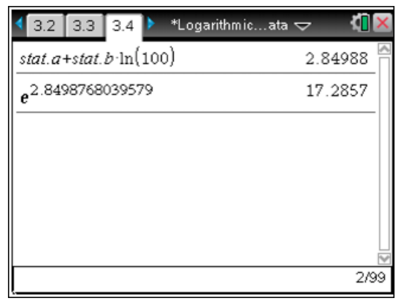
Insert a calculator page, and use the statistics variables generated by the regression function for greater accuracy.

For $a = 100$:
$\ln(\text{period}) = -4.05821 + 1.5007\ln(100) = 2.84988$
$\text{period} = e^{2.84988} = 17.2857$

For a planet with semi-major axis of length 100 AU (astronomical units), the period, or time to revolve around the sun, is approximately 17.2857 earth years.







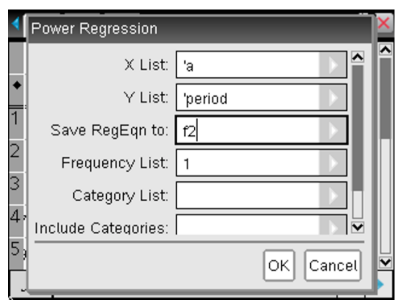
d. Select **MENU > Statistics > Stat Calculations > Power Regression** to find an equation of best fit for the ordered pairs (a, period). Store this equation in $f2$, and add the graph of $y = f2(x)$ to the scatter plot on Page 3.1. Use $f2$ to predict the period for a planet with $a = 100$.

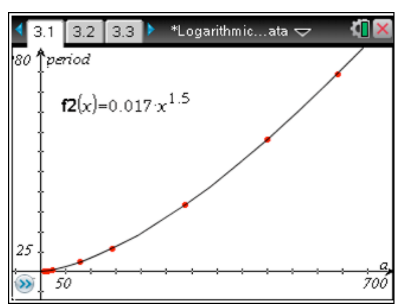
How does this prediction compare with your answer in part (c)?

<u>Answer:</u> The regression equation is
$\text{period} = 0.01728a^{1.50007}$

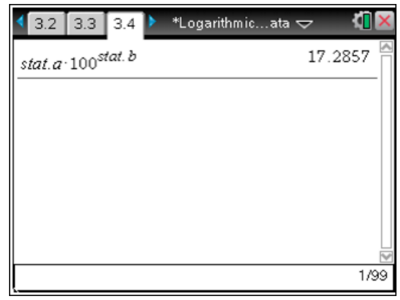
Insert a calculator page, and use the statistics variables generated by the regression function for greater accuracy.

For $a = 100$:
$\text{period} = 0.01728(100)^{1.50007} = 17.2857$





This is the same predicted period as in part (c).



e. Use the power regression equation found in part (d) to solve for period^2, and use the resulting expression to state Kepler's Law of Periods.

Answer: Start with the power regression equation and solve for period^2.

$$\text{period} = 0.017\, a^{1.5} = 0.017\, a^{3/2}$$
$$\text{period}^2 = (0.017)^2\, (a^{3/2})^2$$
$$= (0.000289)\, a^3$$

Kepler's Law of Periods states the square of the period of any planet is proportional to the cube of the semi-major axis of its orbit.

Wrap Up

Upon completion of this activity, students should be able to understand:

- A logarithmic transformation on one or both variables can be used to achieve a linear relationship.
- Linear regression is a technique for finding a line of best fit.
- For exponential regression, logarithmic regression, and power regression, it appears the calculator transforms the data to achieve a linear relationship, and then applies the techniques of linear regression to the transformed data.
- How to use the results from linear regression to convert back to the original relationship between two variables.

TI-Nspire Navigator

Note 1

Question 1a, *Quick Poll (Open Response)*

Tell students that you are going to send a Quick Poll, Open Response asking them to describe the relationship in words.

Note 2

Question 2c, *Screen Capture*

Ask students to guess at the equation of a line of best fit and have them add this graph to Page 1.3. Use Screen Capture to display student responses.

Note 3

Question 2e, *Screen Capture*

Ask students to guess at an exponential equation of best fit before using the calculator function. Ask students to graph their equation on Page 1.2. Use Screen Capture to display student responses.