Transforming Relationships

Student Activity

Name ____________________

Class ____________________

Problem 1 – Analyzing Residual Plots

Run the program **DATA** and select **PART 1**. Press [stat] [enter] to see the data.

The four data sets are: rebound height of a ball dropped from different heights (BOUNCE and HEIGHT), miles per gallon of a vehicle with different weights (MPG and WEIGH), tons of recycled newspaper from 1986–2004 (NEWSP and YEAR), and United States population from 1790–1880 (POP and POPYR).

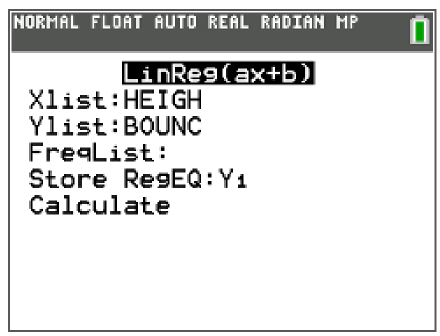
1. Fill in the chart below.

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

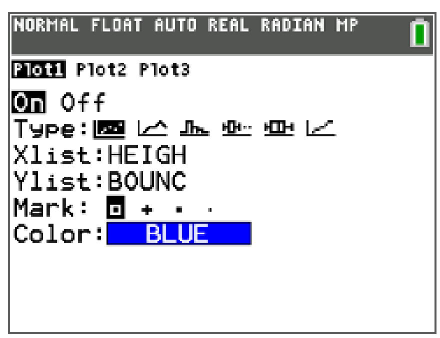
Goal: To analyze the quality of the best fit line for each graph.

Find the linear regression line for the bounce and height graph. To do this, press [stat] and select **LinReg(ax+b)** from the CALC menu.

Select the lists by pressing [2nd] [list]. Press [vars] **> Y-Vars > Function** and select **Y1** from the list.



To view the regression line and graph together, press [2nd] [stat plot] and set up **Plot1** with the setting shown at the right. Press [zoom] and select **ZoomStat**.



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

There are two ways to analyze how well the line fits the data—graphically and numerically.

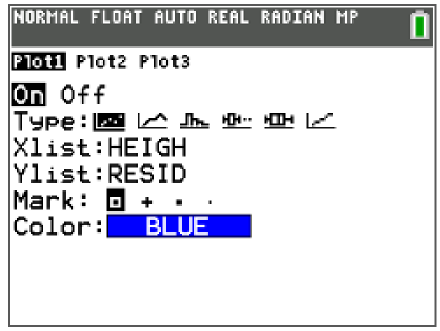
Graphically: Draw the residual plot.

A residual = actual value – predicted value. The residual plot will show the residual for each value of the independent variable. Analyzing the residual plot will allow us to determine if a linear model is the best fit.

A curved pattern shows that the relationship is not linear.

When you found the regression equation, the residuals were calculated and stored automatically in the list named RESID.

Update **Plot1** to the settings shown at the right. Turn off **Y1**.

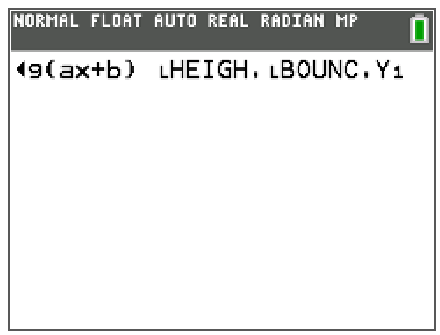


3. Assess the quality of the fit. Explain your reasoning.

Numerically: Calculate the value of *r*, the correlation coefficient.

The closer this number is to 1 or –1, the more linear the relationship is. In addition, r^2 is called the coefficient of determination. It gives the percent of the variation in the dependent variable that can be explained by the linear relationship.

These values can be obtained by turning the diagnostic on and redoing the linear regression as shown at the right.



Press [mode] and select **ON** next to **STAT DIAGNOSTICS**. Press [clear] to exit.

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

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

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

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

	Graphically	**Numerically**
Xlist: WEIGH **Ylist: MPG**		$r =$ $r^2 =$
Xlist: YEAR **Ylist: NEWSP**		$r =$ $r^2 =$
Xlist: POPYR **Ylist: POP**		$r =$ $r^2 =$

Some data have an obvious linear pattern, so a linear model is fitted to the data. Other data have no obvious pattern, so a model is not relevant. Finally, some data have a relationship that is non-linear.

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

Problem 2 – Transforming Data

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

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

Although another type of regression may be a better fit for the data set, note that both r and r^2 are not calculated for all types of regression.

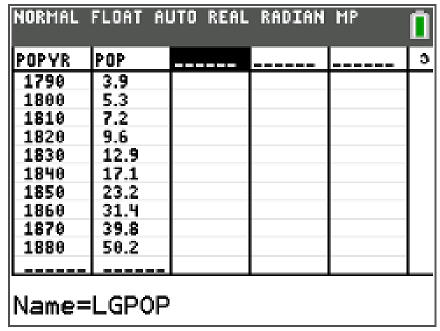
Name ___________________
Class ___________________

Goal: Transform the data to become linear using logarithmic transformation.

Run the **DATA** program and select **PART 2**.

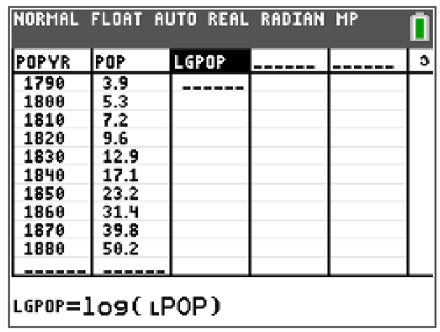
Create a new list in the spreadsheet to compute the logarithm of the population.

With your cursor at the top of POP, arrow to the right, and use the alpha keys to title the new list **LGPOP**.

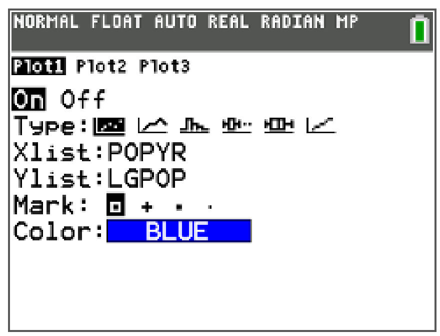


With the cursor at the top of LGPOP, enter **log(LPOP)** and press [enter].

Note: Select the list name **LPOP** by pressing [2nd] [list].



Update **Plot1** to the settings shown at the right. This will graph logarithmic population vs. population year.



Graph **LGPOP vs. POPYR**.

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

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

Graphical:

Numerical: r = ______________, r^2 = __________________.

Name ___________________
Class ___________________

Not all data are linear or exponential, so other types of transformations may need to be performed. If a variable grows exponentially, its logarithm grows linearly. If a power function is used as the model, then a logarithmic transformation of both variables will make the data linear.

Extension – An Additional Transformation

Run the program **DATA** and select **EXTENSION**. This extension uses the lists LAGE, LLENGH, and LWEIGH, which are the average length and weight, respectively, at different ages for Atlantic Ocean rockfish. (The data is from Gordon L. Swartzmann and Stephen P. Kaluzny, *Ecological Simulation Primer*, Macmillan, New York, 197, p. 98.)

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

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

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

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

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