Name ________________________

Class ________________________

Part 1 – Analyzing Residual Plots

The spreadsheet on page 1.3 contains data that will be used in the activity. The four data sets are: rebound height of a ball dropped from different heights, miles per gallon of a vehicle with different weights, tons of recycled newspaper from 1986–2006, and United States population from 1790 – 1880.

BEFORE LOOKING AT THE GRAPHS, 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.

On page 1.4, show the linear regression line for the bounce and height graph. Choose **MENU > Analyze > Regression > Show linear regression (*mx* + *b*)**.

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.***

Choose **MENU > Analyze > Residuals > Show Residual Plot**. A plot of the residuals vs. the independent variable will appear at the bottom of the screen.

1. 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.

Add a calculator page. Choose **MENU >Statistics > Stat Calculations > Linear Regression**. Choose the independent variable for the Xlist and the dependent variable for the Ylist. Choose **OK.**

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

3. How well did a linear model fit the bounce vs. height graph? Explain your reasoning.

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

Now, analyze the three other scatter plots and fill in the chart below.

	Graphically	Numerically
mpg vs. weight		r = r^2 =
newspaper vs. year		r = r^2 =
pop vs. popyear		r = r^2 =

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

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

Part 2 – Transforming Data

6. What type of graph would model the data set you chose in Question 5?

7. Try other regressions from the list under Regression in the Analyze menu. Which do you feel is the best? Why?

Even though another type of regression may fit the data set better, r and r^2 can only be calculated for linear regression.

Goal: Transform the data to become linear using logarithmic transformation

On page 2.1, create a new list in the spreadsheet that takes the logarithm of the population. In column C, title the list “logpop” and enter **=log(pop)** in the shaded box.

On a new *Data & Statistic*s page, graph **logpop vs. popyear**.

8. Does this new graph appear to be linear?

9. 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 = _________________.

Not all data is 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.

In general, power functions are used to transform data into linear models. For example, a square root function or $y = t^{-1}$ can be used to transform data.

Extension – An Additional Transformation

The data in Problem 3 contains the average length and weight 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).

1. Graph **weight vs. length**, add the linear regression line, and calculate the residuals. Assess the fit.

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

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

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

5. What does this tell you about the correct model for the data?