Transforming Relationships

Student Activity

Name ______________________________

Class ______________________________

Lesson Goal: Analyze the appropriateness of a linear regression model for a set of data.

Problem 1 – Analyzing Residual Plots

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

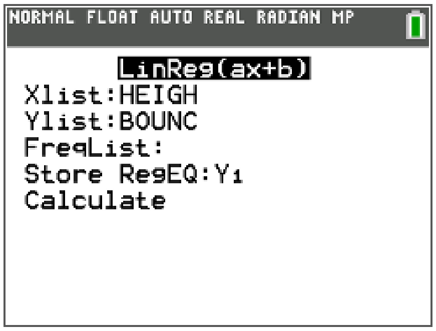
The four data sets are: rebound height of a ball dropped from different heights (BOUNCE and HEIGHT), miles per gallon of a vehicle with different weights (MPG and WEIGH), tons of recycled newspaper from 1986–2004 (NEWSP and YEAR), and United States population from 1790–1880 (POP and POPYR).

1. Identify the independent and dependent variables for each data set.

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

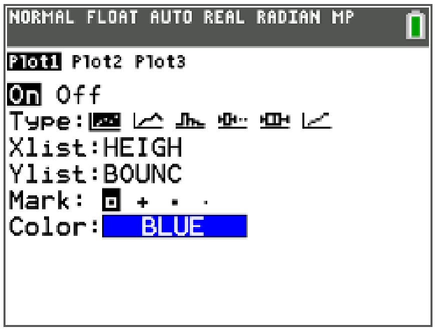
Find the linear regression line for the bounce and height graph. To do this, press stat and select **LinReg(ax+b)** from the CALC menu.

Select the lists by pressing 2nd [list]. Press vars **> Y-Vars > Function** and select **Y_1** from the list to store the regression equation in **Y_1**.



```
NORMAL FLOAT AUTO REAL RADIAN MP
      LinReg(ax+b)
 Xlist:HEIGH
 Ylist:BOUNC
 FreqList:
 Store RegEQ:Y₁
 Calculate
```

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



```
NORMAL FLOAT AUTO REAL RADIAN MP
Plot1 Plot2 Plot3
On Off
Type:
Xlist:HEIGH
Ylist:BOUNC
Mark: □ + · ·
Color: BLUE
```

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

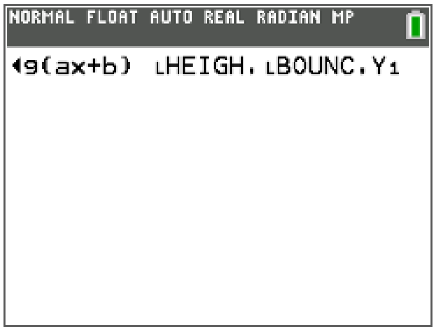
Key Concept: It is **NOT** sufficient to determine the appropriateness of a regression equation simply by visually inspecting the graph. No matter how well a line may *appear* to fit a set of data, there are **TWO** criteria for mathematically determining whether a proposed regression line is an appropriate model for a set of data—numerical and graphical.

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

The closer this number is to 1 or –1, the closer a linear regression is to the data set. The sign indicates whether the relationship is increasing or decreasing. 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 on the diagnostic tool and recalculating the linear regression.

Press mode and select **ON** next to **STAT DIAGNOSTICS**. Press clear to exit. Recalculate the regression as shown at the right.



```
NORMAL FLOAT AUTO REAL RADIAN MP
◂g(ax+b) ʟHEIGH,ʟBOUNC,Y₁
```

3. What are the values of *r* and r^2? What do these values appear to suggest?

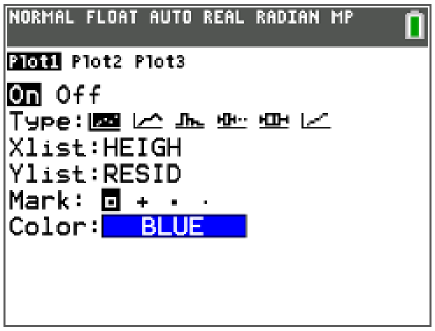
Graphical: Analyze the residual plot.

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.

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**. To see the residual plot, press zoom and select **ZoomStat**.

A healthy residual plot is one in which the points are evenly distributed above and below the *x*-axis, and where there is no clear pattern in plot of points. A curved pattern in the residual plot shows that a linear model is **NOT** appropriate.



```
NORMAL FLOAT AUTO REAL RADIAN MP
Plot1 Plot2 Plot3
On Off
Type:
Xlist:HEIGH
Ylist:RESID
Mark: ▣ + · ·
Color: BLUE
```

4**.** Describe the residual plot. Are the points evenly distributed above and below the *x*-axis? Does there appear to be a pattern in the residual plot, or does it appear to be random?

5. Now that you have analyzed the regression model for **BOUNCE vs. HEIGHT** both numerically and graphically, how appropriate is a linear regression for the data? Explain your reasoning.

6. Use the table below to numerically and graphically analyze the other three data sets.

	Numerically	**Graphically**	**Is a linear model appropriate?**
Xlist: WEIGH **Ylist: MPG**	$r =$ $r^2 =$	Distribution of points relative to *x*-axis: Points randomly dispersed or some obvious pattern:	
Xlist: YEAR **Ylist: NEWSP**	$r =$ $r^2 =$	Distribution of points relative to *x*-axis: Points randomly dispersed or some obvious pattern:	
Xlist: POPYR **Ylist: POP**	$r =$ $r^2 =$	Distribution of points relative to *x*-axis: Points randomly dispersed or some obvious pattern:	

7. Based on the above analysis, which of the data sets appears to have a relationship that is non-linear?