Name: _____ Date: _____

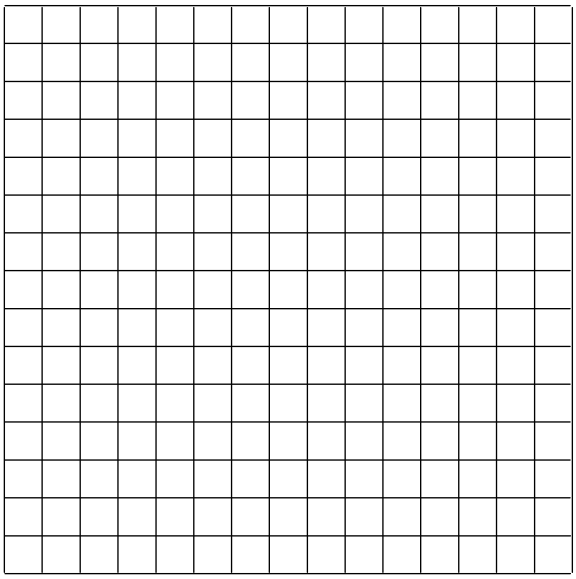
Lines of Best Fit – Linear Regression
Algebra 1

When mathematical models are used in the real world, we often don't have data that fall on perfect lines. Most of the time we want to find the ***line of best fit*** (or the line that best fits the data).

***Exercise* #1:** A biologist is studying the relationship between a tree's diameter and its height. She records the following data for 7 different trees.

Diameter (inches)	2	3	4	5	6	7	8
Height (feet)	8	10	16	17	22	20	29

(a) On the grid provided, create a **scatterplot** of the data. Use the diameter as the independent variable and the height as the dependent variable.



(b) Draw a ***line of best fit*** through the data. As a guide, try to have as many points of data fall above the line as below the line.

(c) Write two ordered pairs that lie on your line.

(d) Determine the equation of your linear function using the two ordered pairs from part (c).

(e) Using your linear function from part (d), estimate, to the nearest foot, the height of a tree given that its diameter is 14 inches. (This type of calculation is called **extrapolating**; we are using a model to predict **outside** of our data set.)

Entering Data in Your Calculator – We will enter data frequently in the calculator. Today we will enter it in order to produce the equation of the line of best fit, which will certainly fit the data better than any rough estimate that we create by hand.

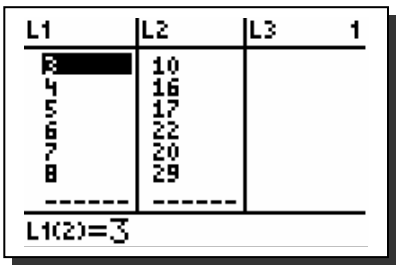
***Exercise* #2:** Consider the same data that you had before:

Diameter (inches)	2	3	4	5	6	7	8
Height (feet)	8	10	16	17	22	20	29

(a) Enter the data into your calculator as follows.

Step 1 – Hit the **STAT** button and go to the **Edit** submenu.

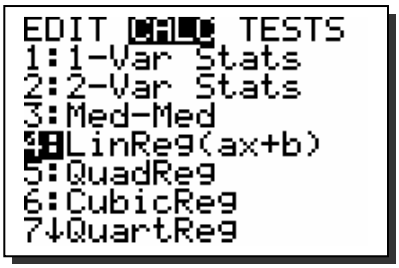
Step 2 – Enter the Diameters under **L1** and the Heights under **L2**. When working with a data set, always place the independent variable in **L1** and the dependent variable in **L2**.

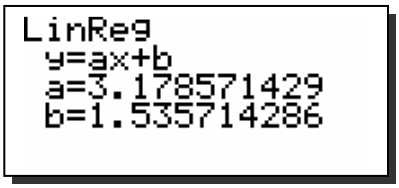


(b) Find the equation for the line of best fit. Round your coefficients to the nearest *tenth*. Also, define what each variable, *x* and *y*, represent.

Step 1 – Hit the **STAT** button and go to the **CALC** submenu.

Step 2 – Go the choice 4 – **LinReg(ax+b)**. Hit **ENTER** twice.





(c) To the *nearest foot*, use your equation to find the height of a tree whose diameter is 14 inches.

(d) Construct a table on your calculator to find the diameter of a tree, to the nearest *tenth* of an inch, for which the tree would be 45 feet tall. Provide numerical evidence to support your answer.

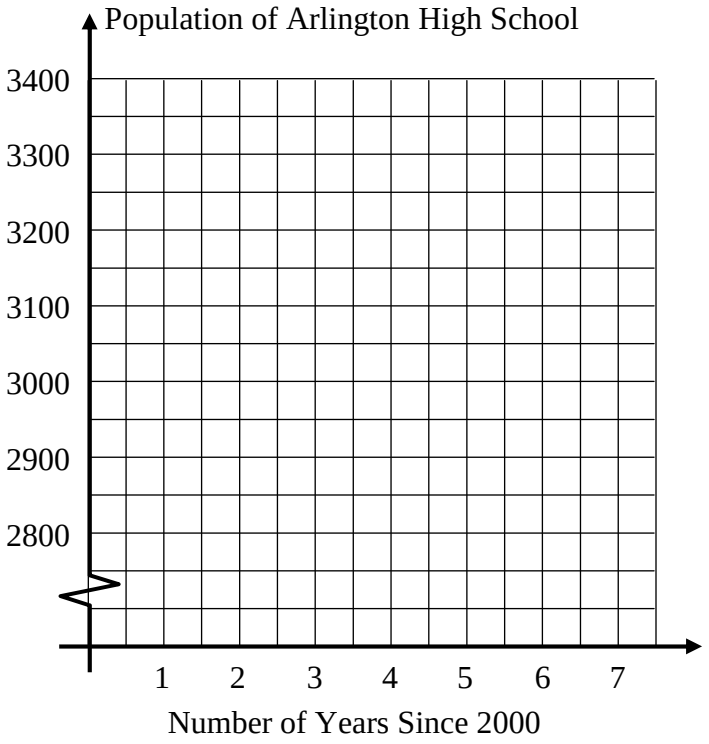
Name: _____ Date: _____

Lines of Best Fit – Linear Regression
Algebra 1 Homework

1. The table below shows the number of students in Arlington High School as a function of the number of years since 2000.

Number of Years Since 2000, x	2	3	4	5	6
Number of Students at Arlington High School, y	2892	3042	3087	3136	3296

(a) On the grid below, draw a scatterplot of the data.



(b) Determine an equation, using your calculator, for the line of best fit. Do not round your coefficients.

(c) Define below what each variable represents:

x represents:

y represents:

(d) Use your model to predict the population of Arlington High School in the year 2012. Round your answer to the nearest whole number.

(e) Use your model and a table on your graphing calculator to determine between which two consecutive whole number years the population reaches 4500. Provide numerical evidence to support your answer.

2. A real-estate agent is trying to determine the relationship between the distance a 3-bedroom home is from New York City and its average selling price. He records data for 6 homes shown below.

Miles from New York City, x	10	35	50	65	75	120
Price of 3 Bedroom Home, y	755,000	650,000	580,000	505,000	475,000	285,000

(a) Using your calculator, write a linear regression equation that relates the distance from New York City, x, to the price of the 3-bedroom home, y. Round your coefficients to the nearest *hundred*.

(b) Woodstock, New York, is located 95 miles from New York City. Using your linear model from part (a), determine the price of a 3-bedroom home in Woodstock.

(c) Using your model, determine the price of a 3-bedroom home in New York City. (Hint: think about the value of x when you are in New York City.)

(d) Using tables, determine, to the nearest mile, the distance from New York City a home would be if its selling price were exactly $500,000. Provide numerical evidence to support your answer.

(e) Using tables, determine, to the nearest mile, the distance from New York City a home would be if its selling price were $0. Provide numerical evidence to support your answer.

(f) Why is your answer to part (e) unreasonable? Explain.

When we use **extrapolation** with linear models we can sometimes get unreasonable answers. This is because we are using the model with independent variable values for which the model does not apply.