Title: Take a B.R.E.A.T.H. with Us!
        (Balloon Regression Experiment Analyzing Time vs. Height) Activity

Brief Overview:

>   The lesson introduces the concepts of line of best-fit (linear regression) and residuals. Data relating the time it takes a balloon to fall from a given height will be collected, a line of best-fit will be graphed, and the residuals will be calculated. A linear regression will be run on the TI-83 to compare the student-predicted equation with the computer-generated result.

Links to NCTM 2000 Standards:

*   **Mathematics as Problem Solving, Reasoning and Proof, Communication, Connections, and Representation**
    These five process standards are threads that integrate throughout the unit, although they may not be specifically addressed in the unit. They emphasize the need to help students develop the processes that are the major means for doing mathematics, thinking about mathematics, understanding mathematics, and communicating mathematics.

    This activity involves the utilization of a variety of strategies applying mathematics to a real-world problem. Class discussions should encourage students to create and explore a variety of mathematical conjectures. Students will be communicating mathematically as they define slope, y-intercept, and best-fit in terms which relate to an investigation. Students also will be introduced to fundamental mathematical concepts while presenting the perfect vehicle for future explorations in physics such as gravity, velocity, and acceleration. Last of all, student generated plots, TI-83 linear regression models, and spaghetti lines of best-fit enable students to more readily visualize the data allowing increased attention to lesson extensions and applications.

*   **Patterns, Functions, and Algebra**
    Students will incorporate slope, y-intercept, domain, and range into the graphical representation of data modeling a linear equation.

*   **Data Analysis, Statistics, and Probability**
    Students will collect data, display it in a scatter plot, and determine a line of best-fit using linear regression. They will calculate the least squares regression. They will be able to explain their own model.

Links to Virginia High School Mathematics Core Learning Units:

*   **A.1**
    The student will solve linear equations and inequalities in one variable, solve literal equations (formulas) for a given variable and apply these skills to solve practical problems. Graphing calculators will be used to confirm algebraic solutions.

- **A.2**

  The student will represent verbal quantitative situations algebraically and evaluate these expressions for given replacement values of the variables. Students will choose an appropriate computational technique, such as mental mathematics, calculator, or paper and pencil.

- **A.5**

  The student will analyze a given set of data for the existence of a pattern, represent the pattern algebraically or graphically, if possible, and determine if the relation is a function.

- **A.6**

  The student will select, justify, and apply an appropriate technique to graph a linear function in two variables. Techniques will include slope-intercept, x- and y-intercepts, graphing by transformation, and the use of the graphing calculator.

- **A.7**

  The student will determine the slope of a line when given an equation of the line, the graph of the line, or two points on the line. Slope will be described as rate of change and will be positive, negative, zero, or undefined. The graphing calculator will be used to investigate the effect of changes in the slope on the graph of the line.

- **A.8**

  The student will write an equation of a line when given the graph of the line, two points on the line, or the slope and a point on the line.

- **A.15**

  The student will determine the domain and range of a relation given a graph or a set of ordered pairs and will identify the relations that are functions.

- **AII.9**

  The student will find the domain, range, zeros and inverse of a function for a given element in its domain, and the composition of multiple functions. Functions will include those that have domains and ranges that are limited and/or discontinuous. The graphing calculator will be used as a tool to assist in investigation of functions, including exponential and logarithmic.

- **AII.19**

  The student will collect and analyze data to make predictions, write equations, and solve practical problems. Graphing calculators will be use to investigate scatter plots to determine the equation for a curve of best-fit.

- **AII/T.8**

  The student will recognize multiple representations of functions (linear, quadratic, absolute value, step, and exponential functions) and convert between a graph, a table, and symbolic form. A transformational approach to graphing will be employed through the use of graphing calculators.

- **AII/T.9**
  The student will find the domain, range, zeros, and inverse of a function; and the composition of multiple functions. Functions will include those that have domains and ranges that are limited and/or discontinuous. The graphing calculator will be used as a tool to assist in investigation of function, including exponential and logarithmic.

- **AII/T.19**
  The student will collect and analyze data to make predictions, write equations, and solve practical problems. Graphing calculators will be used to investigate scatter plots to determine the equation for a curve of best-fit.

- **MA.2**
  The student will perform operations, including composition and inversion of functions, and determine the domain and range of results. Continuity of functions and special functions such as absolute value, step functions, and piece-wise, will be included. Curve sketching and transformations will be included. Graphing utilities will be used. Graphing utilities will be used to investigate and verify the graphs.

Grade/Level:

Grades 7-12; Algebra I, Algebra II, Statistics, AP Statistics

Duration/Length:

5 days (50 minute periods)   A suggested timeline follows the Procedures.

Prerequisite Knowledge:

Students should have working knowledge of the following skills:

- Finding the slope of a line from a graph
- Finding the y-intercept of a line from a graph
- Finding the equation of a line given the slope and the y-intercept
- Finding the domain and range given a set of points or a graph of points

Objectives:

Students will:

- construct a line of best-fit given a set of data points.
- calculate residuals and find their sum.
- decide whether the line is a good fit given the sum of the residuals.
- use a TI-83 calculator to find and graph a linear regression model.
- complete assessment covering objectives.

Materials/Resources/Printed Materials:

- TI-83, one for each student
- Balloons—9-inch diameter minimum
- Stop watch, one for each group
- Measuring stick, or instructor can create tape measures marking 3,4,5,6,7 and 8 feet for each group to be taped to the classroom walls
- Graph paper, preferably with large squares, one for each student
- Color pencils/pens
- Little colored dot labels (quarter inch diameter) from a supply store, 10 for each group
- Spaghetti "line" for each group

Specific materials needed by the instructor:
- Very large (3-4 foot square or use of very large overhead projector) graph paper to be used in front of the class.
- Ten large, brightly colored dot labels
- Magic marker

Development/Procedures:

Step 1:   Class will be divided into groups of three. Each group will be given materials needed and assigned to a particular area of the room.

Step 2:   Each group will divide tasks equally: one timekeeper, one individual who will drop and retrieve the balloon, and one recorder.

Step 3:   Each group will gather data by completing the activities on <u>Worksheet 1</u>. Record data in Table A. Each group's data will be recorded on a class chart and used to average the times for each of the five heights. The teacher may choose to simply choose one group's set of data to save time. Students will record class data in Table B on the worksheet. Consistency of measurement is important; and remind students to consistently measure the height from the bottom of the balloon.

Step 4:   The teacher will guide a class discussion to include: domain, range, how to label each axis, type of scale used for each axis. (x-axis for distance, the independent variable; y-axis for time, the dependent variable.)

Step 5:   Using information gained from the class discussion, the instructor (or a student) will label each axis on the large graph paper; students will do the same on their graph paper.

Step 6:   Students, using the colored quarter-inch labels, will record the data (from Table B) on their graph paper; remind students to label the coordinates of each data point. The instructor will do the same on the large sheet of graph paper.

Step 7:   Students will answer questions #1 - 5 on Activity <u>Worksheet 2</u>.

Step 8:   The instructor will lead a class discussion on recognizing any patterns in the data. (Question #2 on the worksheet.)

Step 9:   Each student, using the spaghetti, will attempt to find the line of best fit for the graphed data. (Remind students that data points are not to be connected with broken spaghetti pieces.) (Question #6 on WS 2)

Step 10:   Students will answer the worksheet questions #7 - 11. Encourage the students to answer the questions in terms of height, time, or average time but not using math terms. Introduce the concept "line of best fit." Show students the symbol for the y-value of each x on the regression line: $\hat{y}$

Step 11:   Students will answer worksheet questions #12 - 14. The instructor will lead the class through a discussion of the strategies they used to find spaghetti lines of best fit, and how one can judge who has the "best" line. The instructor will lead the students to conclude that the "best spaghetti line" means the minimum total distances from the data points to the line.

Step 12:   Each student will start question #15 and Table D on the worksheet. Encourage students to work with their partner. Use the table to record the distances from each data point (**y**-coordinate) to the "spaghetti line of best fit ($\hat{y}$)**.**" Explain to students that this distance is called the **residual**.

Some students will be able to measure the distances visually from the data points to the line of best fit by counting squares on their graph paper; others will need to solve the equations algebraically using the formula and the coordinate points. Students will answer question #16 and total the residual column; remind them not to ignore the "positive" and "negative" distances to the line. (Algebra I and Algebra II students could **use** the absolute value as a stepping stone to understand the **concept** of residuals and that the sum of their distances determines the line of best fit.)

>   **This is a good stopping point for Algebra I students.**
>   **(Assessment may not be appropriate for Algebra I students.)**

Step 13:   The class will then discuss why or why not this is a good measure of fit for their "spaghetti line."   Students should answer question #17. If the students do not notice, the teacher should show them that the sum of the positive and negative residuals is always close to zero. The class usually comes to the conclusion that this is **not** the best way to see whose line is the "winner." The teacher will ask students to propose a different method to judge who has the best line of fit. (Students often suggest absolute value; remind them, however, that statistics **is** calculus-based and that the calculus involved with absolute values is quite complicated.) Students might then suggest squaring the distances.

(Especially when they see the column on their worksheets for squared residuals!) Students should finish the worksheet by answering questions #18 and #19.

This is a good stopping point for Algebra II students.
(Assessment Worksheet – Page 1 Only)

Step 14: The class will be formally introduced to the concept of **residual** as the vertical distance from the data point to the line of best fit. Students will also be introduced to the new vocabulary, symbols, and formula for residual: ($y - \hat{y}$).

Step 15: The class will discuss what they have found and the instructor will introduce the Sum of Squares formula-- $\sum (y - \hat{y})^2$. Now determine which student pair chose the line with the smallest sum of squares.

Step 16: Using Worksheets 3 - 5 and the TI-83, students will record the data in lists, graph a scatter plot of the data, find the line of best fit using linear regression (option #4; for AP Statistics, use #8) and find residuals. Instructions will be on the worksheets.

Step 17: Class discussion on the findings of each group.

Step 18: Student Assessment Worksheet for both Statistics and AP Statistics classes.

Suggested timeline:
  Day 1: Data collection (steps 1 – 4)
  Day 2: Graphing a line of best-fit (steps 5 – 9)
  Day 3: Analysis of line of best-fit (steps 10 – 13)
  Day 4: TI-83 calculator model and wrap-up (steps 14 – 17)
  Day 5: Assessment (step 18)

Assessment:

The teacher should conduct an on-going assessment of student progress by circulating among cooperative student groups during class activities, and by evaluating student contributions to class discussion and questions. The teacher may also grade the activity sheets. A final student assessment worksheet with scoring rubric is provided.

Students will demonstrate that they know how to find a "line of best fit" on the calculator. The student will demonstrate an understanding of the process that determines "best fit" by finding the sums of squares for two different lines using the same set of data and correctly deciding which one is the line of best fit.

Possible teacher rubric for evaluation

1. Student finds equation of one line.
2. Student finds equations of both lines
3. Student fills out one table accurately and completely.
4. Student fills out both tables completely and accurately.
5. Student mentions that the best fit line is determined by the lowest sum of squares.
6. Student finds the regression line correctly with the TI-83

Extension/Follow Up:

1. Conduct the activity using a different object, such as a feather. Discussion topics will include comparing the slopes of the linear regression equations from both activities.

2. Use the model for prediction.

3. A project could be used as an optional assessment tool. Students would find published data and present a mathematical model of the data using linear regression. A possible web site could include: www.lib.stat.cmu.edu.

4. Quadratic modeling.

5. Use a beaker, marbles and water to model height of water when marbles are added.

6. Use a beaker with water and a small boat to model height of boat using a CBR to collect the data.

Authors:

Lisa Dawson
Loudoun Valley High School
Loudoun County,VA

Cathy Furlong
Langley High School
Fairfax County, VA

Terry French
Lake Braddock Secondary School
Fairfax County, VA

Susan Hrizdos
Wheeling Park High School
Ohio County, WVA

Take a B.R.E.A.T.H. With Us Activity Worksheet 1
(**B**alloon **R**egression **E**xperiment **A**nalyzing **T**ime vs. **H**eight)

I.   Introduction

Students will gather data by dropping a 9-inch balloon from various distances, develop a regression model using spaghetti and a TI-83 graphing calculator, and calculate residuals.

II.   Materials

Each group will need the following materials:

1 stopwatch	1 9-inch (or larger) balloon
1 strip of colored dots	1 piece of uncooked spaghetti for each student
1 tape measure	1 TI-83 graphing calculator for each student
1 student worksheet	

III.   Procedure

(A)   Students form groups of at least 3 members. One student will be the timekeeper, one student will drop and retrieve the balloon, and one student will record the data gathered.

(B)   Students will hold a balloon in such a way that the bottom of it is held at a distance 3 feet above the floor. The balloon will then be dropped and a student (using a stopwatch) will record the time it takes the balloon to reach the floor. Record this time in the first row of Table A.

(C)   Students will then drop the balloon from a new location (1 foot higher than the last) and record the time it takes to reach the floor. Record this time in Table A.

(D)   Repeat step (B) 4 more times.

(E)   In order to calculate the class average, the teacher will collect data student data. These values will be placed into Table B.

(F)   Plot the class data by assigning each value in the first column of Table B (distance) to x, and the corresponding entry in the second column (time) to y. Use Graph C and the colored dots to plot these 6 data points.
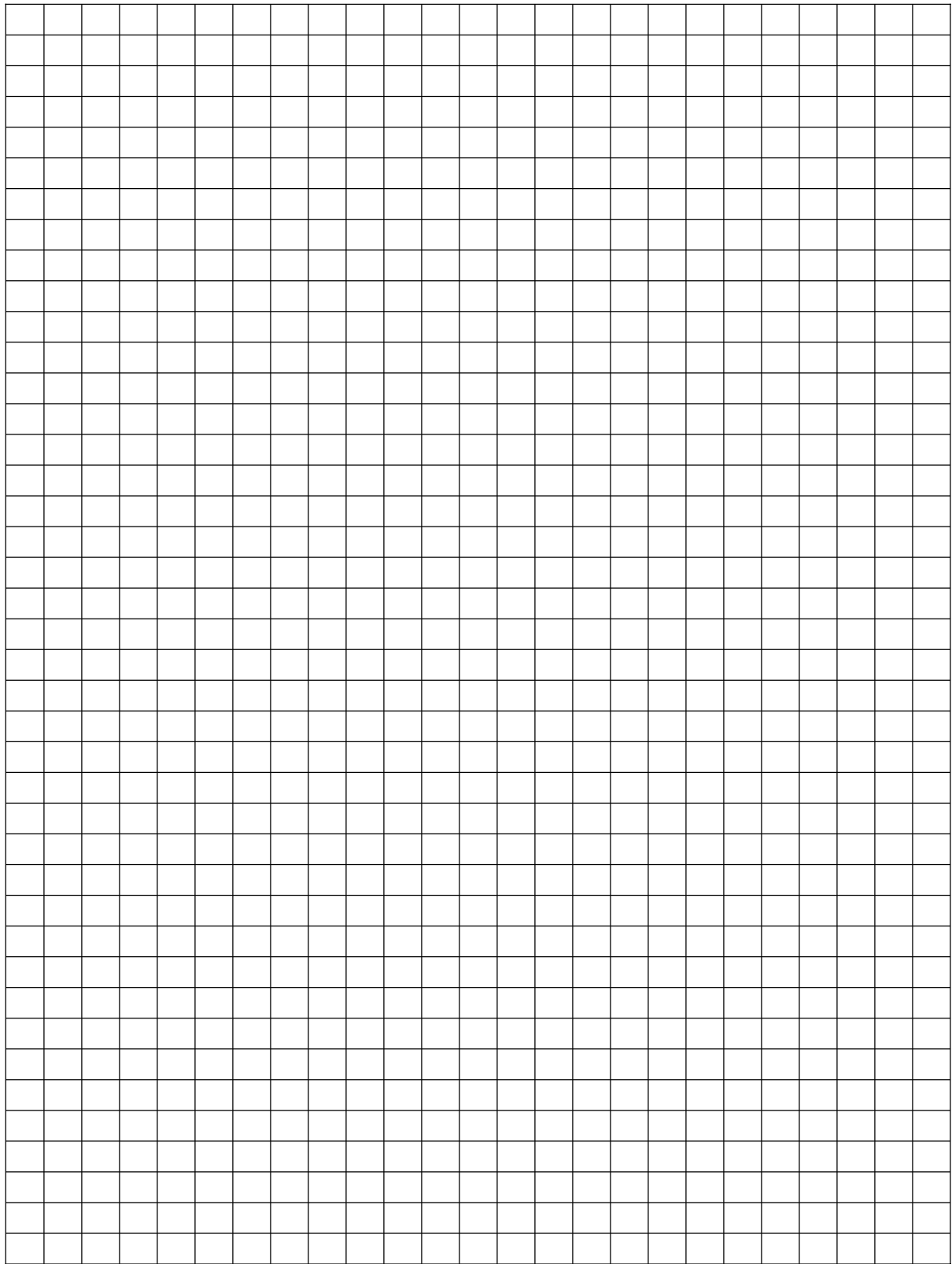
Group Data

Distance (in feet) x	Time (in seconds)
3	
4	
5	
6	
7	
8	

Table A

Class Data

Distance (in feet) x	Time (in seconds) y
3	
4	
5	
6	
7	
8	

Table B

Graph C

Take a B.R.E.A.T.H. With Us Activity Worksheet 2

(1)   In your own words, describe the relationship between the distance and time.

(2)   Do you see any patterns in the scatter plot of class data?
       If so, what?

(3)   Does this data represent a function?

(4)   What is the domain of this function (the x-values or distance)?

(5)   What is the range of this function (the y-values or time)?

(6)   Use a piece of spaghetti to find a line that **best** describes or *fits* your data points. Trace this line onto your graph. This is your "spaghetti line".

(7)   Choosing any 2 points on your spaghetti line of best fit, calculate the slope (m). Use the formula:

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$ to find the slope.

       Slope = _____

(8)   In your own words, explain what the slope measures in this investigation.

(9)   Find the y-intercept (b) of your spaghetti line of best fit.
       y-intercept = _____

(10)   In your own words, explain what the y-intercept measures in this investigation.

(11)   The slope-intercept form of a line is $\hat{y} = \mathbf{mx} + \mathbf{b}$. Write the slope-intercept equation of your spaghetti line of best fit.

Slope-intercept equation _____

(12)   What strategy did you use to position the piece of spaghetti through your data points in Graph C?

(13)   Justify why you think the piece of spaghetti models your data points as a line of best fit.

(14)   How far away from the spaghetti line of best fit is each data point?

(15)   Keeping in mind that the spaghetti line of best fit has a slope of m and a y-intercept of b, complete Table D. Let $\hat{y}$ represent the y-value for each x-value on the line.

Table D

Distance (x)	Observed Time (y)	Spaghetti line $\hat{y}$ = mx + b	Residual y - $\hat{y}$	Square of Residual $(y - \hat{y})^2$
3				
4				
5				
6				
7				
8				
		Totals		

(16)   Find the sum of the Residual column in Table D.

Sum of residuals = _____

(17)   Is this a good measure of closeness to the spaghetti line of best fit? _____ Explain.

(18)   Find the sum of the Squares of Residuals column in Table D.

Sum of squared residuals = _____

(19)   Is this a good measure of closeness to the spaghetti line of best fit? _____
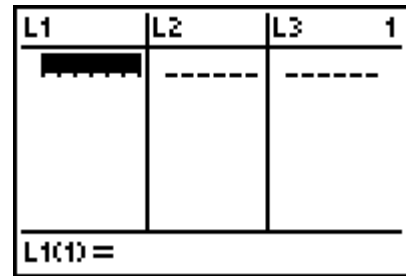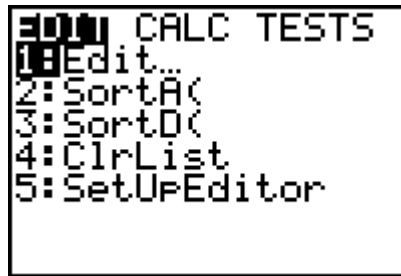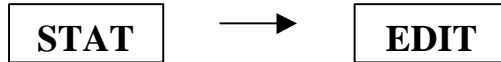Explain.

Worksheet 2 – Answers

(1)   As distance increases, the time also increases.

(2)   The data appears to cluster about a straight line with positive slope.

(3)   Yes. Each element in the domain is paired with exactly one element in the range.

(4)   The x-coordinates.

(5)   The y-coordinates.

(7)   Slope = Answers will vary.

(8)   The time it takes the balloon to fall a given distance. (rise/run)

(9)   y-intercept = Answers will vary.

(10)   The y-intercept measures the time needed for the balloon to fall when it's already on the ground.

(11)   Answers will vary.

(12)   Answers will vary.

(13)   Answers will vary.

(14)   Answers will vary.

(15)   Table of values will vary.

(16)   Answers will vary.

(17)   No, because there are some positive and some negative values. These tend to cancel each other out so that the sum is very close to zero.

(18)   Answers will vary.

(19)   Yes, because if there is a lot of distance from the points to the line, the value will be bigger than if there is just a little bit of distance form the points to the line. So this measures what we want.
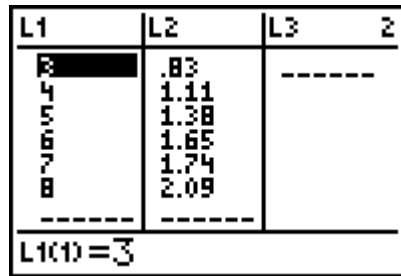
Take a B.R.E.A.T.H. With Us Activity Worksheet 3

Graphing a Scatter Plot of the Data on the TI-83

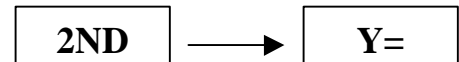(A)   Enter the data from Table A into the calculator's Statistics Editor. Press the following keys:

$$\boxed{\text{STAT}} \quad \longrightarrow \quad \boxed{\text{EDIT}}$$

```
EDIT CALC TESTS
1:Edit…
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor
```

```
L1      L2      L3      1
▄▄▄▄▄   ------  ------


L1(1)=
```

Now enter the **distances** into L_1 and the corresponding **times** into L_2.

```
L1      L2      L3      2
3       .83     ------
4       1.11
5       1.38
6       1.65
7       1.74
8       2.09
------  ------
L1(1)=3
```

(example)

(B)   Set up a plot by pressing the following keys: $\boxed{\text{2ND}} \longrightarrow \boxed{\text{Y=}}$

```
STAT PLOTS
1:Plot1…Off
    L1      L3      □
2:Plot2…Off
    REIGN 1         □
3:Plot3…Off
    L6      1
4↓PlotsOff
```

```
Plot1  Plot2  Plot3
On  Off
Type: ▨  ⟋  ▥
      ⊹  ⊞  ⟋
Xlist:L1
Ylist:L2
Mark: □  ■  ·
```

Select Plot 1.   **Turn Plot 1 on.**
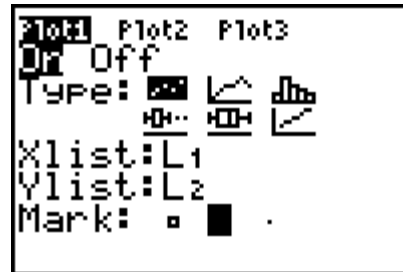   **Select the options as shown**
                                                                above.

(C)   Set up the viewing window. Look at your data in Table A. Decide upon values of x and y that will represent the investigation's results appropriately. Complete Table E.

(D)   Press   **WINDOW**   and enter the values from Table E into the TI-83's viewing window.

Window Settings	Values
Xmin	
Xmax	
Xscl	
Ymin	
Ymax	
Yscl	

Table E

(E)   Graph the data points by pressing   **GRAPH**   .

(F)   Explore the values of each data point by pressing   **TRACE**

OR

The TI-83 will set a window for you by   **ZOOM**   →   **9**

Take a B.R.E.A.T.H. With Us Activity Worksheet 4

Finding the Line of Best Fit on the TI-83

(A) Go to the home screen: | 2nd | → | QUIT |

Calculate the least squares line: | STAT | → | CALC | → | 4 |

```
EDIT CALC TESTS
1:Edit…
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor
```

```
EDIT CALC TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
```

Type the x-list and the y-list: | 2nd | → | 1 | → | , | → | 2nd | → | 2 |

To graph the line onto the scatter plot: | , | → | VARS | → | Y-VARS | → | 1 | → | 1 |

```
VARS Y-VARS
1:Window…
2:Zoom…
3:GDB…
4:Picture…
5:Statistics…
6:Table…
7:String…
```

```
VARS Y-VARS
1:Function…
2:Parametric…
3:Polar…
4:On/Off…
```

```
FUNCTION
1:Y₁
2:Y₂
3:Y₃
4:Y₄
5:Y₅
6:Y₆
7↓Y₇
```

```
LinReg(ax+b) L₁,
L₂,Y₁
```

Press   ENTER

(B)   Record the values from the linear regression screen in the table below:

LinReg
   y=ax+b
   a=
   **b=**
   r^2=
   r=

(C)   Write the equation of the line from your calculator:_____

NAME_____PERIOD_____DATE_____

Take a B.R.E.A.T.H. With Us Activity Worksheet 5

Calculating Residuals Using the TI-83

(A)   How does your spaghetti line of best fit compare to the calculator's line of best fit? Is it
       exactly the same?   Is it close?

(B)   You've already calculated the residuals and the sum of squares for your spaghetti line, so now
       see how well you did by calculating the same information for the calculator's line. The TI-83
       stores the list of residuals for each data point after a regression is run. To find this list and
       to store it in L_3:

| 2nd | → | STAT | → | ↓ | → | RESID | → | ENTER | → | STO→ | → | 2nd | → | 3 |
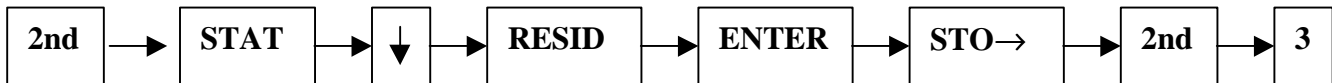
```
NAMES OPS MATH
1:A
2:ACTUA
3:AP
4:APFRQ
5:DF
6:GUESS
7↓HT
```
```
NAMES OPS MATH
 ↑PSQM
 :REIGN
 ■RESID
 :WT
 :X
 :Y
 :YEAR
```
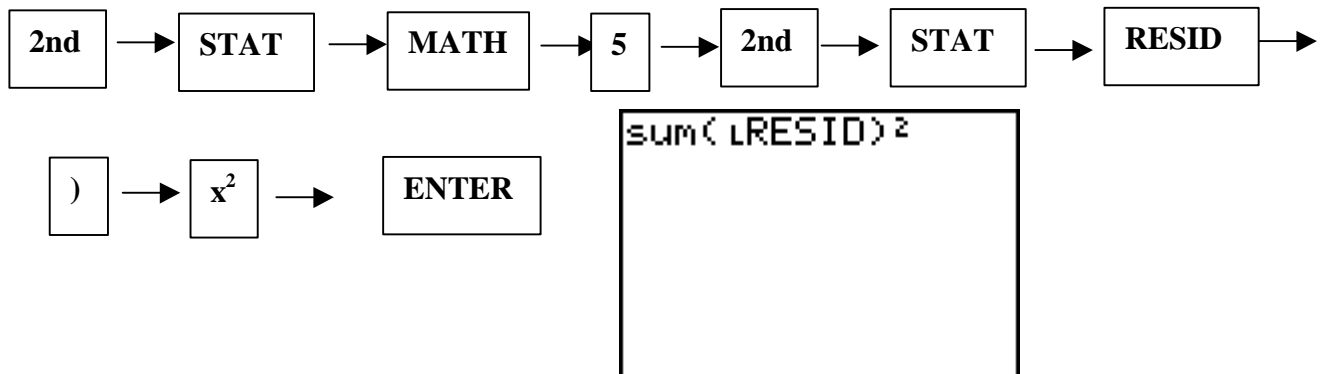```
LRESID→L3■
```

Press:    ENTER

The data is now stored in $L_3$.  To see it, go to the Stat Edit screen. (See part A at the beginning
of  WS 3.)  To find the sum of the squared residuals on the home screen (it can also be done in
the data window)

| 2nd | → | STAT | → | MATH | → | 5 | → | 2nd | → | STAT | → | RESID | → |

| ) | → | $x^2$ | → | ENTER |

```
sum(LRESID)²
```

(C) What value do you get after you pressed ENTER?_____

(D) Is this smaller than the sum of squares you calculated for your spaghetti line?_____

(E) Which is the better line (line of best fit)?_____ Why?

NAME_____PERIOD_____DATE_____

Student Assessment

The following scatter plots represent the same data points but have different "lines of best fit". Completely answer the following questions, then choose which line best represents the data. Your answer must be justified both mathematically and logically.

Scatter Plot I   Scatter Plot II

POUNDS   POUNDS

WEEKS   WEEKS
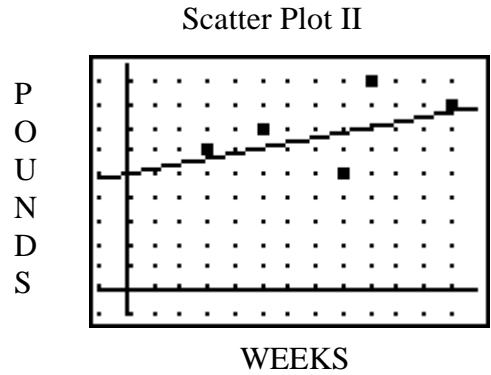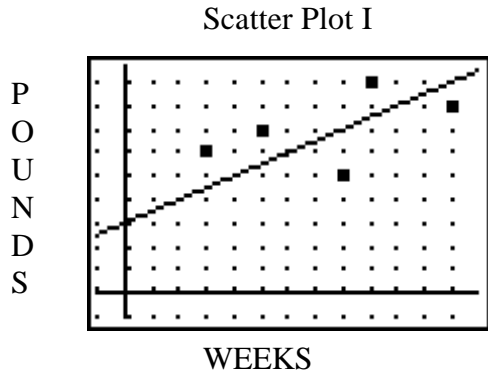
Both scatter plots have a scale of one.

Complete the following questions on the above data.

1. Equation of the line of best fit for Scatter Plot I: _____

2. Equation of the line of best fit for Scatter Plot II: _____

3. Complete the table using information from Scatter Plot I

WEEKS (x)	POUNDS (y)	POINT ON THE LINE $\hat{y} = mx + b$	RESIDUAL $y - \hat{y}$	SQUARE OF RESIDUAL $(y - \hat{y})^2$
3				
5				
8				
9				
12				
			Total	

4. Complete the table using information from Scatter Plot II

WEEKS (x)	POUNDS (y)	POINT ON LINE $\hat{y} = mx + b$	RESIDUAL $y - \hat{y}$	SQUARE OF RESIDUAL $(y - \hat{y})^2$
3				
5				
8				
9				
12				
			Total	

5.   Choose which line best represents the data. Justify your answer using information from questions #3 and #4 ***only.***

6.   Use the TI-83 to calculate the regression line for this data. $\hat{y} =$ _____

7.   Which equation, #1 or #2, is closest to the equation in #6?_____

Student Assessment - ANSWER KEY

1. **y = 0. 5x + 3**   2. **y = 2/9x + 5**   (Both of these can be obtained by counting on the graphs. Answers may vary somewhat.))

3.

WEEKS (x)	POUNDS (y)	POINT ON THE LINE $\hat{y}$ = mx + b	RESIDUAL $y - \hat{y}$	SQUARE OF RESIDUAL $(y - \hat{y})^2$
3	6	4.5	1.5	2.25
5	7	5.5	1.5	2.25
8	5	7	-2	4
9	9	7.5	1.5	2.25
12	8	9	-1	1
			TOTAL	11.75

4.

WEEKS (x)	POUNDS (y)	POINT ON LINE $\hat{y}$ = mx + b	RESIDUAL $y - \hat{y}$	SQUARE OF RESIDUAL $(y - \hat{y})^2$
3	6	5.7	.3	.09
5	7	6.1	.9	.81
8	5	6.8	-1.8	3.24
9	9	7	2	4
12	8	7.7	.3	.09
			TOTAL	8.23

5. **y = 2/9x + 5** is the line of best fit for the given data. Students should justify the answer by comparing differences in the Sums of the Squares. The equation with the lower number is best. So, since 8.23 is smaller than 11.75, the second line is best.

6. y = 0.22x + 5.3 **OR** for AP Statistics classes: y= 5.3 + 0.22x

7. The second equation is closer to the regression line from the TI-83. 2/9 = .2222... and the 5 is very close to 5.3.