# Drawing a Regression Line between

# Spag & B

*Students make a slam dunk, first with spaghetti and a scatter plot, then with a graphing calculator.*

S. Aslı Özgün-Koca and Thomas G. Edwards

STUART BURFORD/ISTOCKPHOTO.COM

# hetti
# asketball

*"I liked to grab things on the graphics and see it first hand . . . see what we were doing on the problem. It makes more sense."*

These words were from a student who completed a series of lessons on lines of best fit. She made this comment after comparing a hands-on approach using a strand of spaghetti with a high-tech approach using a calculator.

To give students meaningful experiences while studying regression lines, we designed a learning environment that accommodates *Bruner's three forms of representation* (1966, pp. 10–11). Bruner advocated a sequence of learning that runs the gamut from using physical materials to using symbols to teach mathematical ideas effectively:

1. The *enactive form of representation* refers to physical actions being performed on objects. When students have physical interactions with manipulatives, kinesthetic experiences result.

2. With *iconic representation,* images instead of physical objects are put to work. Direct manipulation of objects is replaced by the beginning of more abstract thinking with visual or other perceptual imagery.

3. *Symbolic representation* involves a manipulation of symbols using a formal system of syntactic rules. Students' abstract reasoning continues and develops in conjunction with symbols at this stage.
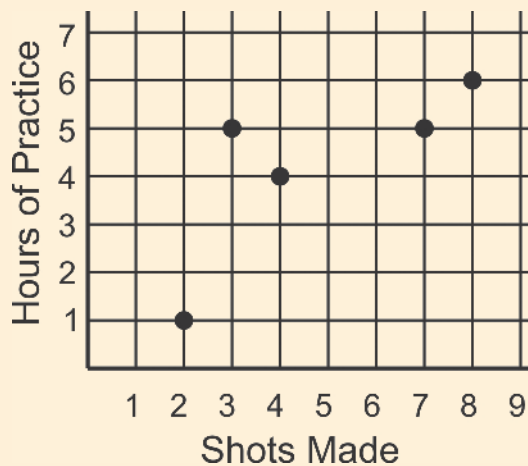
We follow Bruner's route with an activity that will be familiar to many readers (see, for example, Mathematics TEKS Refinement 2006). We continue our manipulation through technology, and end the activity with symbol usage. On one hand, we believe, as do Beck and Huse (2007), that

advantages [of virtual manipulatives] include a linkage from the specific to the general, the possibility of posing a wider variety of problem types, the establishment of a strong foundation for problem solving, an increased focus of attention and motivation, and an encouragement of accurate student explanations for mathematics strands. (p. 482)

Mr. Chemberler is a middle school basketball coach. He wants to study the relationship between the number of hours per week that a player practices shooting foul shots and the number of successful foul shots made during a team competition. He collected the following data for his five starting players.

| Hrs. per Week of Practice | Foul Shots Made (Out of 10) |
|---|---|
| 3 | 5 |
| 8 | 6 |
| 4 | 4 |
| 2 | 1 |
| 7 | 5 |



on a graphing calculator by rotating and moving the line and observing the results of these movements on the error, which was automatically calculated by the calculator.

## A REGRESSION ACTIVITY USING SPAGHETTI

We asked students to consider the problem situation and the plotted points in the graph provided in **figure 1**. While they determined a line of best fit using a piece of spaghetti taped to the graph, we asked:

- Could the line pass through all the points?
- Could the line pass through some of the points?
- Could the line not pass through any of the points?

This graph was presented to students without labels and scale marks on the axis. Our students said that the line could not pass through all the points because the points are scattered, but it can pass through some of them. They also agreed that the line need not pass through any of the points.

To encourage students to compare different possibilities, we asked the following:

Suppose two students, Jack and Jill, have two different lines represented by their spaghetti strands. How could we help them decide which line fits the data better?

Our students agreed that the better fitting line should be closer to the data points. When we asked how to measure how close, they suggested using a ruler to find how far the points were from the line. We then negotiated measuring the vertical distance using the grid on the graph paper as a way to standardize the measurements.

Students then investigated how good a fit their lines were by measuring

Perl (1990), on the other hand, highlights the need for both types of experiences—physical and virtual—and the need to balance their use: "Both software and manipulatives have a role in improving math learning and . . . the two together are more powerful than either alone" (Perl 1990, p. 21).

A major premise in developing these activities is the importance of both physical and virtual manipulation of mathematical concepts and a specific sequence of activities. The primary goals were to introduce the concept of linear regression with a hands-on activity centered on a real-world problem and use technology to focus more on the concepts of the line of fit, error, and the correlation

coefficient. Although lines of best fit are often taught in eighth grade, many teachers and students might not have investigated the concept of the correlation coefficient. Therefore, this activity is intended for the upper middle school grades and lower high school grades. We used the activity with eighth graders.

We began with an activity in which students manipulate an uncooked spaghetti strand on a scatter plot to determine the best-fit line. Once they eyeballed a best-fit line, they calculated the error by measuring the vertical distance, that is, the residual, of each data point from their line. Next, we moved this activity into a technological environment where students manipulated their line of fit
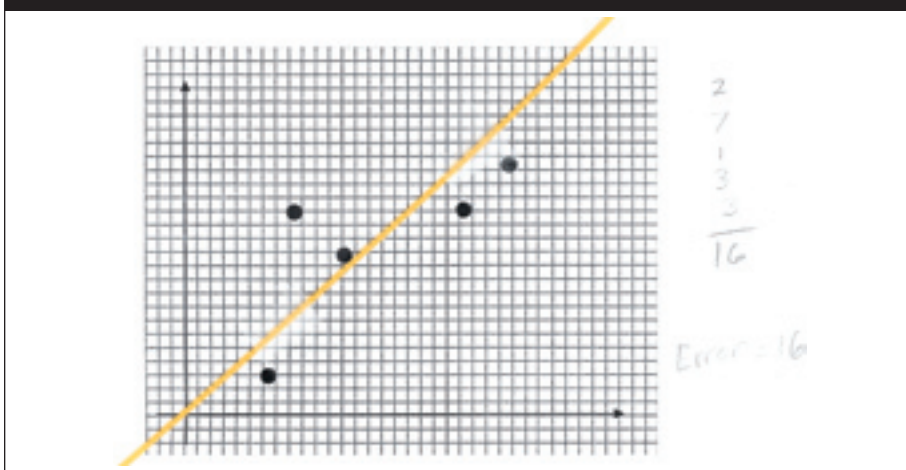
the vertical distance from each point to their spaghetti line and determined the total error by summing all the distances (see **fig. 2**). Because the students were actively engaged in finding the distance from a point to a line geometrically by counting units, the issue of order never arose. They simply counted the units. Whether they began at the point or at the line, it was intuitively obvious to them that the distance should be positive. This scenario would be a good segue to absolute value, because the algebraic representation of counting the units is subtracting the *y*-coordinates.

**Figure 2** displays a student's calculated error of 16 after counting the vertical distance from each data point to the line. However, the data points were plotted on a coordinate axis where three small squares formed a unit. Since most students were focused on creating the smallest error by manipulating the spaghetti strand by hand, many forgot to divide the calculated error by three after counting small squares. Thus, the student's calculated error of 16 was actually $16 \div 3 = 5\ 1/3$ in the context of the graph. However, at this point, students were more interested in obtaining the smallest error. Finally, students compared their results to see whose line best fit the data and discussed why.

## REGRESSION ACTIVITY WITH TI-NSPIRE

Students began the technology-based activity by opening a pre-existing document in the TI-Nspire™ calculator. Although teachers can create this file with students during instruction, we suggest completing and downloading it to the students' calculators beforehand to conserve valuable class time. Students saw that the same data points were plotted, a virtual spaghetti strand was placed on the same axes, and the error was calculated; that is, the "error" is the total of the vertical



**Fig. 2** A student used spaghetti to trace the regression line and calculated the error by hand.

distances from each point to the line. At this point, students saw the vertical lines drawn from the data points to the line and were asked to interpret "error = 7.13936."

They easily connected the calculator's "error" to that calculated during their work with a spaghetti strand. Next, they explored whether that error could be made smaller by replacing, rotating, and moving their virtual piece of spaghetti. At this point, students were focusing on what 7.13936 meant and how to make this number smaller or larger as they manipulated the virtual spaghetti (see **fig. 3a**).

They used the cursor tool to point to the graph until the cursor changed to a double arrow; they then clicked and held the hand button. The cursor changed to a hand, which looked like it was holding the line. They could then change the slope of the graph and observe the changes in the symbolic form and the error (see **fig. 3b**).

Their job now was to make the error as small as possible. To move their virtual spaghetti up and down, they pointed to the middle of the line. The cursor changed to four arrows. They clicked and held the hand button to grab the line and move it up or down, which allowed them to continue to observe the changes in the symbolic



**Fig. 3** Manipulating the virtual spaghetti on the TI-Nspire allowed students to see, in graphic form, the error of their line of best fit.

(a)

(b)

(c)
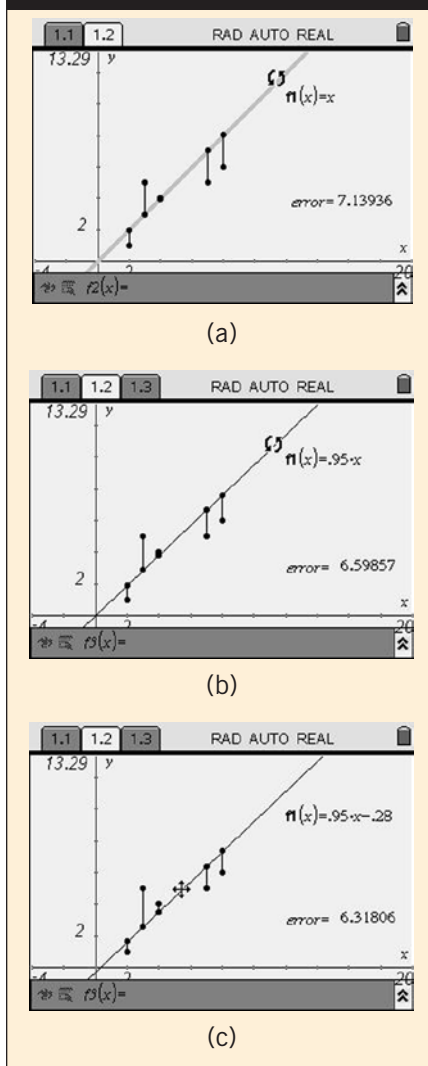
form and the error. An even smaller error resulted (see **fig. 3c**). These questions were asked:

- Were you able to get an even smaller error?
- What is the equation of the line that produced the smallest error?

A lively competition ensued among students to find the smallest error. Students manipulated their virtual spaghetti by rotating it and moving it up and down. As students obtained (and shouted out) their ever-smaller errors, a new goal was created for the rest of the class. This competitive aspect occurred much more prominently in the virtual activity, possibly because of the immediate feedback from the calculator regarding the size of the error when students were moving the virtual spaghetti.

At this point, we wanted students to compare the physical activity with the calculator activity:

- Is the calculator's best-fit line the same as your spaghetti line?
- If not, how are they different?
- If they are different, how is the equation of the calculator's best-fit line different from the equation of your best-fit line?
- What do the slope and $y$-intercept mean as you think about Mr. Chemberler's players?

We also inquired about the correlation coefficient, $r$. When students executed linear regression in their calculators, they saw that $r$ was calculated automatically as .76 (see **fig. 4a**). The correlation coefficient described the relationship between two variables. The variables were the number of hours of practice per week and the number of foul shots made out of 10 attempts. Because it was positive, large values of one variable tended to be paired with large values of the other, and small values of one variable tended to be paired with small values of the other. A correlation of .76 was considered moderately strong; a perfect positive correlation would have been 1, and no correlation at all would have been 0.

Since these students had previously studied the concept of correlation, they had the experience and knowledge necessary to have a sound discussion about $r$. One student wrote: "A good correlation is closer to 1." For students without a previous knowledge of correlation, this could be a great opportunity to start discussing the idea of correlation coefficient, which describes the direction and the magnitude of a relationship. In addition, students can discuss what this correlation coefficient means within the context of the problem: The more hours per week that one practices, the more foul shots out of 10 that a player tends to make.

Finally, we wanted students to consider the effects of changing one or more points in the scatter plot



**Fig. 4** As students changed the data points, they saw the effect on the line of best fit and the correlation coefficient.

(a)

(b)

(c)
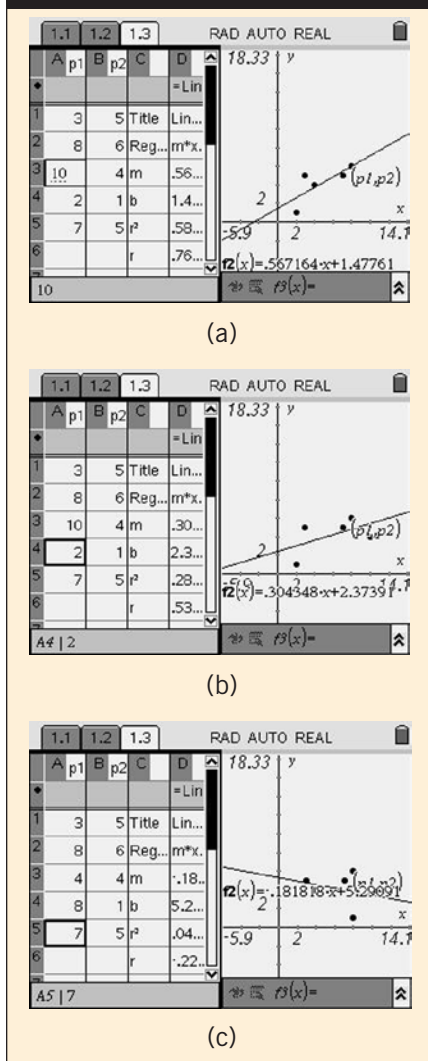
(see **fig. 4**). After moving to the data window by hitting the Ctrl and Tab buttons, students changed the points to study what occurred to the best-fit line. To change the coordinates of a point, they clicked on a cell and typed the new number. As we made changes in the coordinates of one or more points in the data set, students were asked, "How do the best-fit line and the correlation coefficient, $r$, change?"

Because our students had previous experience with the idea of correlation, they easily related a strong correlation between $x$ and $y$ to a value of $r$ that was close to ±1. One student summed it up:

"For a better relationship, you want to get it closer to 1." However, some had difficulty associating a weak correlation with values of *r* that were close to 0. We asked them to think of correlation as a line segment with −1 at one endpoint and 1 at the other endpoint.

"If ±1 represented a strong correlation, where would a weak correlation fall on that line segment?" we asked.

"At zero," they said.

During this part of the activity, we replaced (2, 1) in the original set of points with (8, 1) (see **fig. 4c**), which led to a "wow!" moment:

*Student 1:* Whoa!
*Several students:* Now it's negative!
*Teacher:* How could you tell that so fast?
*Student 2:* Now it [the line] is going down.
*Teacher:* What property of the line changed to negative?
*Student 3:* The slope!

We were happy to see eighth graders make a connection between the correlation coefficient and the slope of the best-fit line.

## STUDENTS' THOUGHTS

During a class discussion, we asked students to compare the two activities. An overwhelming majority preferred manipulating the virtual spaghetti instead of the real spaghetti. They also liked the linkages between representations and the dynamic feature of the technological environment.

Students liked being able to see "the error right on the screen, so it was easier to find the line of best fit" as they were moving the line. Some students mentioned that they liked having exact, accurate, quick, and effortless calculations. "[You] move it [the line] really little and you see the exact error" and "I liked manipulating the virtual spaghetti because it was easier and more precise. You didn't

have to measure the virtual spaghetti, either. The calculator did it for us."

The students raised another point about the links provided by the technology concerning the effects of changing data points on the correlation coefficient: "It showed me that one data point could change it [the correlation coefficient] dramatically." Some thought that the new calculator was confusing and was different from those they had already used, but most also said that with time and experience, the new version would be no more difficult to use than the previous version.

Few students showed their preference for manipulating the real spaghetti. Some mentioned that they liked the simplicity and being able to touch the best-fit line. One student commented about learning the meaning of the error during the hands-on experience; another student thought that manipulating the real spaghetti was not helpful "because you can move the spaghetti one point and you really wouldn't know how much better that was than the first point."

## CONCLUSION

These activities have shown how employing the power of a state-of-the-art calculator can enrich a well-known activity that uses a physical manipulative to support the development of a mathematical concept. In doing so, we believe that the calculator-based virtual manipulative serves as a link between the concrete physical manipulative and the abstract mathematical concept in its symbolic representation.

Bruner's enactive, iconic, and symbolic forms of representation were employed with this activity:

1. Students had a hands-on experience when they manipulated the spaghetti-strand line kinesthetically (an enactive representation).
2. Students then used a calculator to

manipulate the graph of a best-fit line (the digital images were iconic representations).
3. Students considered and interpreted the equations of the best-fit lines, as well as the error and correlation coefficient displayed by the calculator (symbolic representations).

The payoff for students was that they experienced two linked levels of support for an abstract concept. We believe that is the basis for Perl's contention that, although physical and virtual manipulatives are both valuable in their own right, together they can create more powerful learning environments than either by itself.

## REFERENCES

Beck, Shari A., and Vanessa E. Huse. "A Virtual Spin on the Teaching of Probability." *Teaching Children Mathematics* 13 (May 2007): 482–86.

Bruner, Jerome S. *Toward a Theory of Instruction.* New York: W. W. Norton, 1966.

Mathematics TEKS Refinement 2006. "Foundations for Functions." Stephenville, TX: Tarleton State University, 2006. http://www.tarleton.edu/~team/MTR%20HS/lessons/algebra1_7.pdf.

Perl, Teri. "Manipulatives and the Computer: A Powerful Partnership for Learners of All Ages." *Classroom Computer Learning* 10 (1990): 20–22, 26–29.

**S. Aslı Özgün-Koca,** aokoca@wayne.edu, teaches mathematics and secondary mathematics education courses at Wayne State University in Detroit, Michigan. She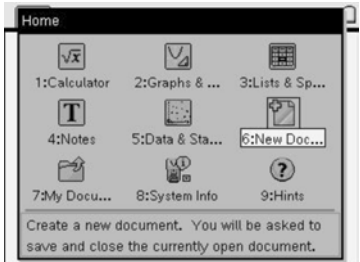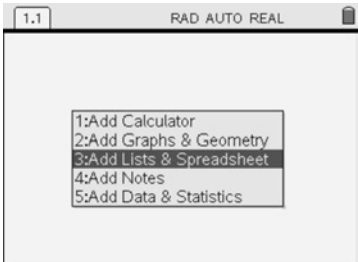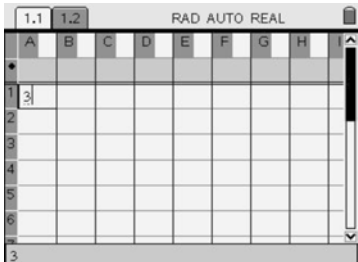 is intere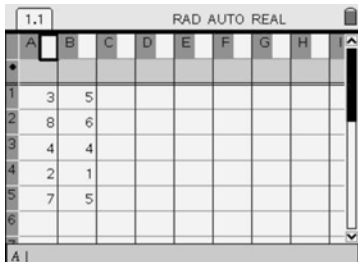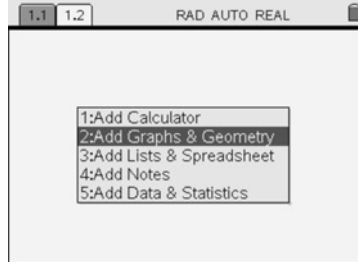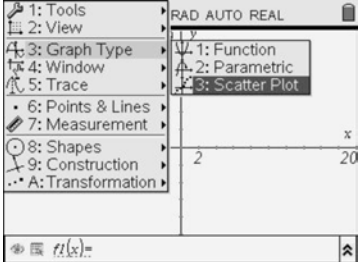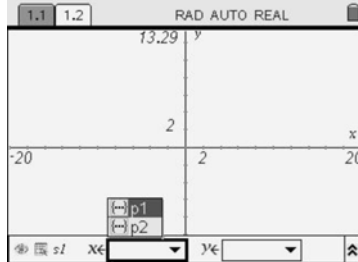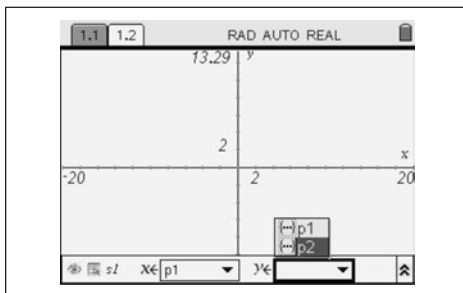sted in the use of technology in mathematics instruction. **Thomas G. Edwards**, t.g.edwards@wayne.edu, works in the College of Education at Wayne State University. He has been interested in the use of graphing calculators to enhance mathematics instruction since their invention.
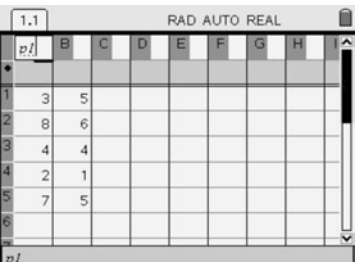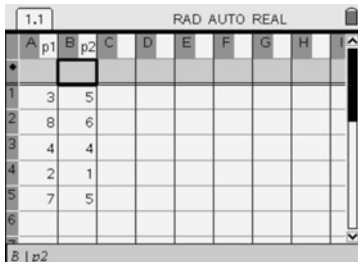
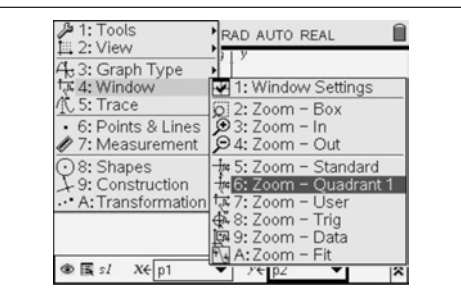# Drawing a Regression Line between Spaghetti and Basketball

## CREATING WITH THE TI-NSPIRE

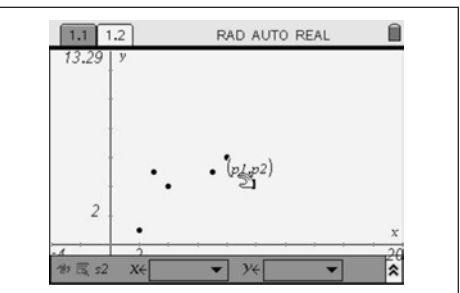To create a document on the TI-NSpire, walk through these steps.
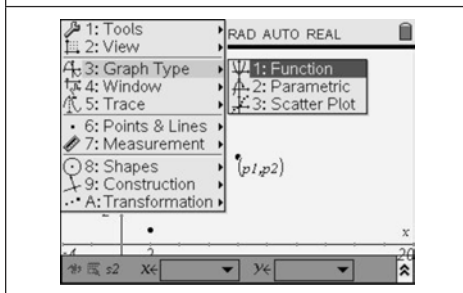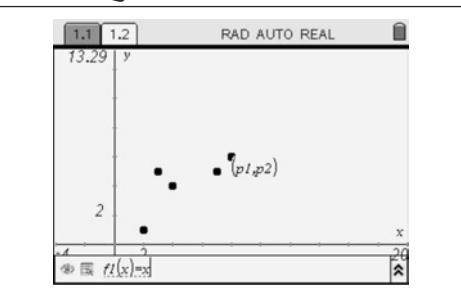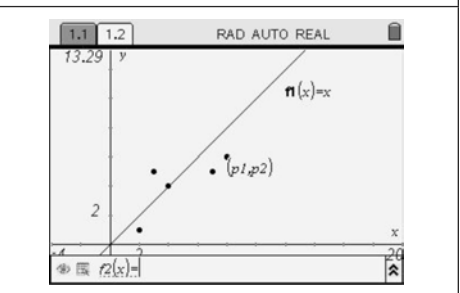
| | | |
|---|---|---|
|  |  |  |
| 1. Hit the Home button on the calculator and then select "6: New Document." | 2. When prompted to add a new page, select "3: Add Lists and Spreadsheets." | 3. The spreadsheet is displayed. Enter data points. |
|  |  |  |
| 4. After entering all data points, name the column by going up to the cell next to the letter *A* by using the arrow keys. | 5. Name column A as *p1* by typing *p1* and hitting Enter. | 6. Name column B as *p2* by typing *p2* and hitting Enter. |
|  |  |  |
| 7. Start a new page by hitting the Ctrl button and then the letter *i*. When prompted, choose the option "2: Add Graphs and Geometry." | 8. Change the graph type to scatter plot by hitting Menu and selecting the option "3: Graph Type" and then "3: Scatter Plot." | 9. To enter *p1* in the *x* box, hit Enter when the *x* box is highlighted, choose *p1*, and hit Enter again. |

| | | |
|---|---|---|
| <br><br>10. Use the right arrow key to highlight the *y* box, hit Enter, and select *p2* using the arrow keys. Hit Enter again. | <br><br>11. To see the data points in the first quadrant more closely, hit Menu and then option "4: Window" and "6: Zoom: Quadrant 1." | <br><br>12. The zoomed graph is displayed. |
| <br><br>13. To place a strand of virtual spaghetti on the graph, change the graph type to function by hitting Menu, selecting "3: Graph Type" and "1: Function." | <br><br>14. Type *x* at the Entry line for *f1*(*x*). | <br><br>15. The graph of *f*(*x*) = *x* will appear on the screen after hitting Enter. |
| <br><br>16. To measure the distances of each data point to the line, create line segments by hitting Menu, "6: Points and Lines," and "5: Segment." | <br><br>17. Point to a data point. It will start flashing. Hit Enter to start the line segment. | <br><br>18. Continue your segment to the graph of the line *f*(*x*) = *x*. Point to the line so that it will start flashing. Hit Enter again to end the line segment at the line. |
| <br><br>19. Repeat the same process for each of the other data points. (See the next two screens for (4, 4).) | <br><br>20. For the data point (4, 4), move the line upward. Hit the tab once to go to the entry line, scroll up to *f1*(*x*), and type +1 next to *f1*(*x*)=*x*. | <br><br>21. The line will move up. To create a line segment for the data point (4, 4), hit tab twice to go back to the graph screen. |

| | | |
|---|---|---|
| 22. To measure those line segments, hit Menu and select option "7: Measurement" and then "1: Length." | 23. Point to a line segment to see the measurement. Small differences could occur in the measurements. | 24. Hit Enter and choose the measurement place. Hit Enter again. |
| 25. After repeating these steps for each line segment, start the process of adding all of those distances. Hit Menu and select "1: Tools" and then "5: Text" to write the formula. | 26. After selecting a spot for the formula and hitting Enter, a box will appear. | 27. Write your formula as $x+y+z+t+u$, and hit Enter. |
| 28. Calculate the formula; hit Menu and select "1: Tools" and "7: Calculate." | 29. Point to the formula to calculate, and hit Enter. | 30. The calculator will ask for the variables $x$, $y$, $z$, $t$, and $u$. Point to each of the measurements of the line segments, and hit Enter in any order. |
| 31. Once each variable is selected and defined, the calculation will appear. Carry that result to any point, and hit Enter. | 32. To clean up the screen, hide some of the calculations by hitting Menu and selecting "1: Tools" and "2: Hide/Show." | 33. Point to the items and hit Enter. After selecting all items to hide, hit the esc button. |

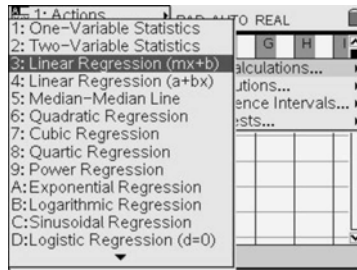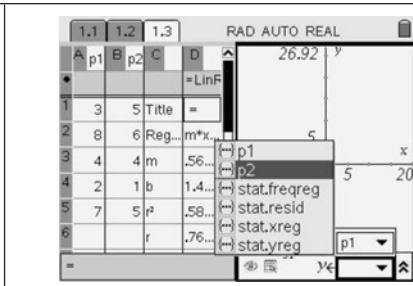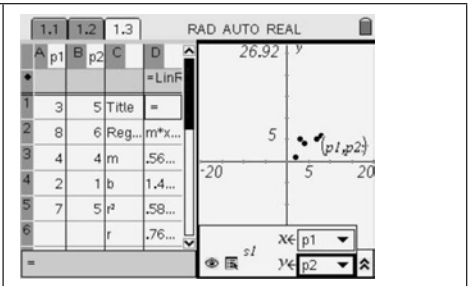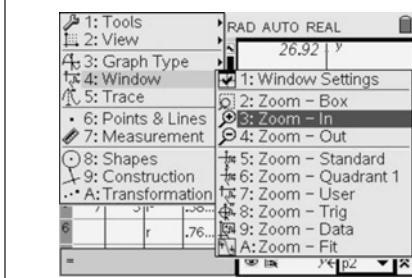| | | |
|---|---|---|
| 34. The screen will be cleaned. | 35. Label the calculation of the sum by adding a text box (see how to add text, above). | 36. To go back to *f1*(*x*)=*x*, hit tab once to go to the entry line, scroll up to *f1*(*x*), type *x* for *f1*(*x*). |
| 37. A new error will appear. Again, small differences could occur in the calculations. | 38. To calculate the best-fit line, create a new page by hitting the Ctrl button and the letter *i*. | 39. Select "3: Add Lists and Spread-sheets." Type *p1* for column A, *p2* for column B, and hit Enter to see the data points. |
| 40. Next hit Menu and select "4: Sta-tistics" and "1: Stat Calculations." | 41. Select "3: Linear Regression (mx+b)." | 42. Choose *p1* for X list, hit tab to move to Y list, and choose *p2* for Y List. Hit tab until the OK button is highlighted, and hit Enter. |
| 43. When prompted, choose Variable Reference. | 44. The results of the analysis will show up on the next two columns. | 45. For a split screen, select wrench, with the Ctrl and Home buttons. Select "6: Page Layout," "2: Select Layout," then "2: Layout 2." |

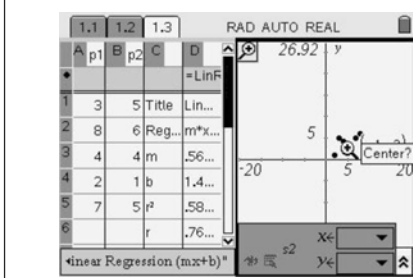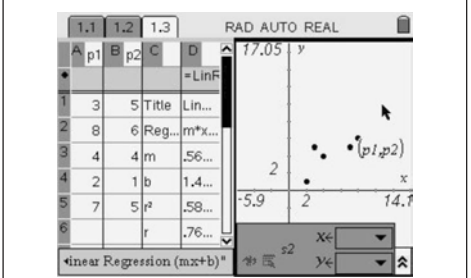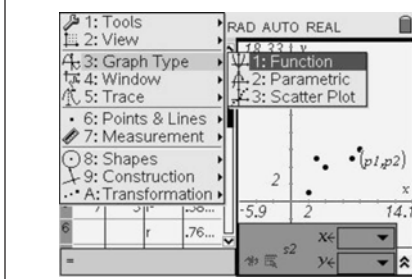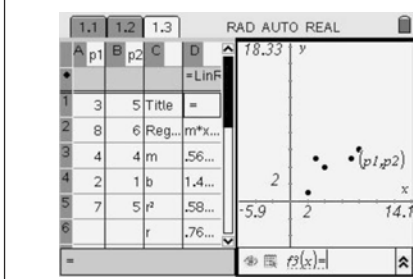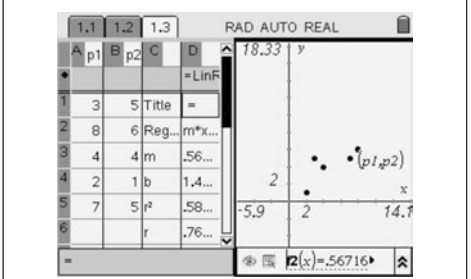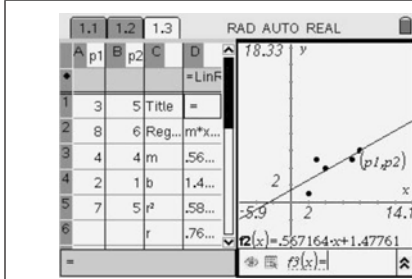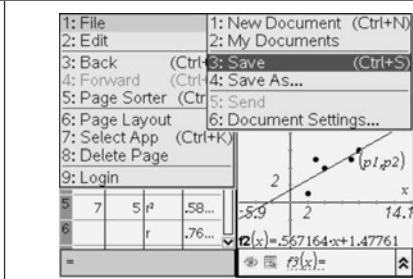| | | |
|---|---|---|
|  |  |  |
| 46. Hit the Ctrl and Tab buttons to move between the two screens. After hitting Menu and selecting graph application, create a scatter plot on the right-hand side of the screen (see above). | 47. Hit Enter to see possible data points to enter in the x box, and choose *p1* for the X List. Hit Enter again. | 48. Use the right arrow key to high-light the y box. Hit Enter, and choose *p2* for the Y List. Hit Enter again. |
|  |  |  |
| 49. Zoom into the data points by hit-ting Menu, and select "4: Window" and "3: Zoom-In." | 50. Select a center for zoom. | 51. The graph can be seen much more closely. |
|  |  |  |
| 52. To see the graph of the best-fit line on the same axes, change the graph type back to function by hitting Menu, "3: Graph Type," and "1: Function." | 53. Nothing is graphed at this point. | 54. Scroll up to *f2(x)* (which holds the Regression Equation) in the Entry Line, and hit Enter. |
|  |  | |
| 55. The graph of the best-fit line will appear. | 56. Save the document by hitting the wrench button (Ctrl and Home), choosing "1: File," then "3: Save." Choose a name such as "spa" (short for spaghetti). | |