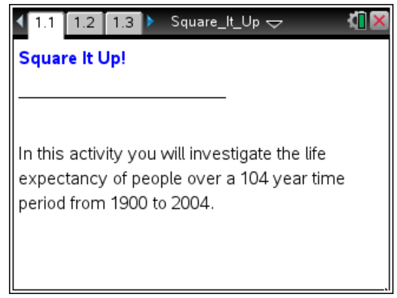


Math Objectives

- Students will represent data on a scatter plot, and they will analyze the relationship between the variables.
- Students will fit a linear function to a scatter plot and analyze the fit by plotting residuals.
- Students will model data with mathematical equations and use appropriate technology tools strategically.

Vocabulary

- scatter plot
- linear regression
- least squares regression
- residual

About the Lesson

- In this activity, students will draw a scatter plot and a movable line. They will investigate the method of least squares by adding the squares to the scatter plot and moving the line to find the minimum sum. Students will then compare their line to the built-in linear regression model.
- This activity is intended to be **student-centered**. The worksheet is designed for students to work independently, and then answer a set of inquiry questions with a partner.
- Prior to this lesson, students should be able to construct a scatter plot.
- The solutions TI-Nspire document shows the scatter plot of each set of data (men, women, and person), as well as the regression line for each.

TI-Nspire™ Navigator™ System

- Send out the *Square_It_Up.tns* file.
- Monitor student progress using Class Capture.
- Use Live Presenter to spotlight student answers.

Activity Materials

- Compatible TI Technologies: TI-Nspire™ CX Handhelds, TI-Nspire™ Apps for iPad®, TI-Nspire™ Software

Tech Tips:

- This activity includes screen captures taken from the TI-Nspire CX handheld. It is also appropriate for use with the TI-Nspire family of products including TI-Nspire software and TI-Nspire App. Slight variations to these directions may be required if using other technologies besides the handheld.
- Watch for additional Tech Tips throughout the activity for the specific technology you are using.
- Access free tutorials at http://education.ti.com/calculators/pd/US/Online-Learning/Tutorials

Lesson Files:

Student Activity

- Square_It_Up_Student.pdf
- Square_It_Up_Student.doc

TI-Nspire document

- Square_It_Up.tns

Discussion Points and Possible Answers

The goal of this exploration is to see how the calculator finds the line of best fit, known as the linear regression model. It uses a technique called "Least Squares Regression."

This activity will allow students to plot data, draw a line on the graph with the corresponding squares, and then transform the line to minimize the sum of the area of the squares. Then they will compare their line with the regression model.

Move to page 1.2

The spreadsheet on page 1.2 displays the life expectancy for men, women, and people in the United States for every ten years from 1900 to 2000 and each year from 2001 to 2004. (Source: http://www.cdc.gov/nchs/data/hus/2010/022.pdf.) Students will investigate the relationship between the year and the life expectancy for a person.

Move to page 1.3

On this Data & Statistics page, students will create a scatter plot to illustrate the relationship between the **year** and the life expectancy for a **person** by selecting **year** on the horizontal axis and **person** on the vertical axis.

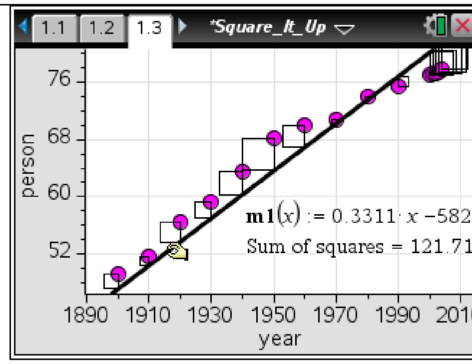
1. Describe the relationship between the two variables.
   Answer: As the years increase, the life expectancy for a person increases. The relationship may be linear.

Add a movable line to the graph and move it to estimate the line of best fit.

> **Tech Tip:** Press menu > **Analyze > Add Movable Line**. Adjust the line until you feel that it best fits your data by using the following steps:
> Move the cursor to the middle of the line until cross arrows (✣) appear.
> Press ctrl [grab] to grab the line and a closed hand will be displayed.
> Move the hand vertically to shift the line up and down.
> When the line is in the correct position, press esc to release the hand.
> Move the cursor to either end of the line to change the slope of the line.
> When the curved arrows appear (⟲), grab the line with ctrl [grab].
> Note that the equation updates as you move the line.

> **Tech Tip:** Tap [tool] and choose **Analyze** > **Add Movable Line**. Grab and move the line near the middle of the line to change the *y*-intercept. Grab and move the line near the end of the line to rotate it.

In the scatter plot, **year** is the independent variable on the horizontal axis and **person** is the dependent variable on the vertical axis. Students should see that the relationship of the variables is almost linear.



Teacher Tip: When students first add the movable line and the residuals, some of the squares will be larger than the others. Students should understand that the larger the square, the greater the residual.

Analyze the residuals.

From the menu, choose **Analyze > Residuals > Show Residual Squares**. A square will appear at each data point. The side of the square is equal to the value of the residual.

Residual = Actual value – Predicted value

2. If a data point is above the regression line, is the value of the residual positive or negative?

   Answer: When a point is above the line the residual value is positive.

3. If a data point is below the regression line, is the value of the residual positive or negative?

   Answer: When a data point is below the line, the residual value is negative.

Find a least squares regression line

A **least squares regression line** is found by minimizing the sum of the area of the squares. Have students move the line to minimize the size of the squares. Students will transform the movable line to try to *minimize the sum of the squares* by adjusting the *y*-intercept or the slope.

4. What is the smallest value you can find for the **sum of squares**?

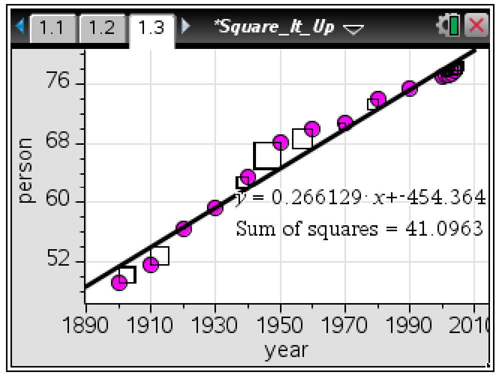
   Answer: Answers will vary, but as they will see with the regression line, the smallest value possible is 41.0963.

5. How many of the points are above your line? How many points are below your line?

   Answer: Of the 15 points, about 7 are above and 7 are below.

6. Check with a partner. How does his/her value compare to yours? If it is different, explain how.

   Answer: The values should be similar, but not identical. Students should record the differences among the sum of the squares, slope, and *y*-intercept.



Students are asked to compare their sum with a partner's. Then, they are given a set of questions that direct them to observe the following properties of a line of best fit.

i. There should be an approximately equal number of points above and below the line.

ii. The points should be distributed throughout the line; there should not be clusters of points above (or below) the line in one area.

iii. The points on the residual plot should not form a pattern. Students will not see the residual plot until later in the activity.

In general, the smaller the sum of the squares gets, the closer the line becomes to the linear regression line. As it approaches the line, the three properties should become more evident.

7. Which data points increase the sum of squares the most? Where do these occur in the data?

   Answer: Answer will vary, but students should recognize that the data points that increase the sum of the squares are the data points which are farthest from the line. The 1950 data point appears to increase the sum of the squares the most. These points are in the middle of the data set.

8. How does the size of these squares compare to the others?

   Answer: Estimating from the graph, the 1950 data point of 68 years of age appears to have a positive residual of about 4. So, that would contribute 16 to the sum of the squares value, which is 41.

9. What do you notice about the distribution of the data points around the line (i.e., above vs. below the line, or equal spacing vs. clusters)?

   Answer: About half of the data points are above and half are below. The years 2000 to 2004 are clustered together, and the residuals before 1950 get larger while they get smaller after 1950.

10. To have the smallest sum, how do the points need to be distributed?

    Answer: To have the smallest sum the data points need to be near the equation model.

Use technology to calculate a least squares line of best fit.

On the same scatter plot, students will compare their line with the one generated by the device. To do this, they should select **Show Linear (mx+b)**. Comparing the placement of the line and the sum of squares will give students a better understanding of how a regression should fit the points.

11. What is the value of the sum of squares?

    Answer: 41.0963.

12. How does the regression line differ from your line? How much did the slope differ?

    Sample Answer: Answers will vary, but students should note the differences between the slopes and *y*-intercepts.

13. How well did you do? Compare your sum of squares with the one generated from the regression line.

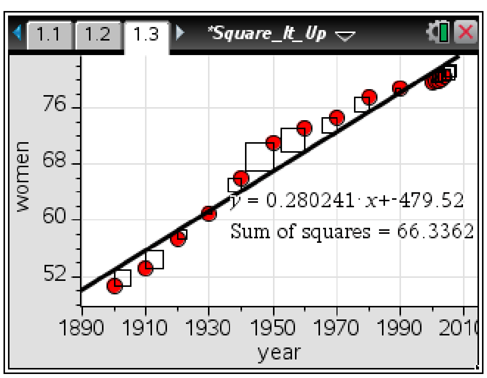
    Answer: Answers will vary. Ideally, students will obtain a sum of squares that is within the range 41.0963 to 42.0963. If students obtained an answer less 41.0963, there is something wrong.

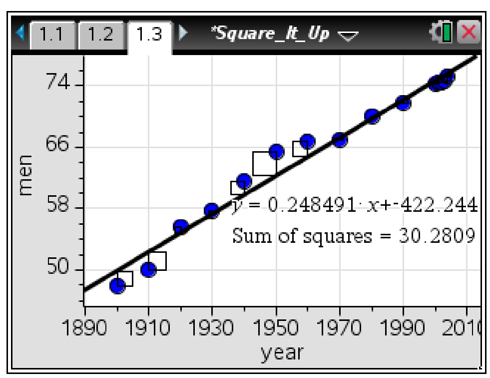
14. What advice would you give to a friend who was trying to find the least squares regression line? What techniques would work best for him/her? (Do not answer with "trial and error.")

    Sample Answer: Adjust the slope so that it approximates the slope of the data. Then move the line until the sum of the squares is as small as possible. See the three properties outlined before question 7. There should be the same number of points above and below the line.

Extension

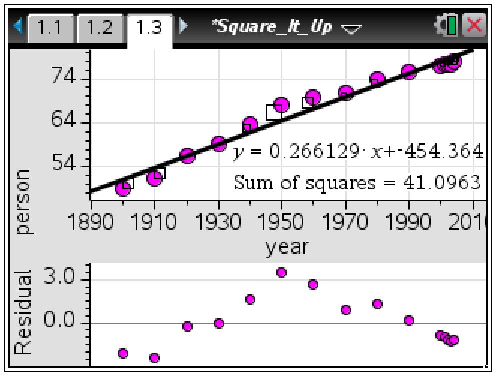
Students can analyze the two additional data sets and complete the table on the student worksheet. Have students discuss what is similar and different among the three scatter plots and regression lines. Sample screenshots and answers are given below.





	Women	*Men*
Your equation	Answers will vary	Answers will vary
Sum of squares	Answers will vary	Answers will vary
Calculator equation	$y = 0.280241x - 479.52$	$y = 0.248491x - 422.244$
Sum of squares	66.3362	30.2809

Further discussion about the data set can take place with the extension. A residual plot can be drawn on the same screen on which the data is plotted. To do so, select **Menu > Analyze > Residuals > Show Residual Plot**.



The residual plot shows a pattern that suggests that the data are not linear. A more appropriate model of this data can be investigated.

A different discussion that can take place involves using the regression line to predict values.

15. Use the linear regression model to predict the life expectancy in 2010. What would be the life expectancy in 2020?

    Answer: Using $y = 0.266129x - 454.364$, when $x = 2010$ the life expectancy is 80.56. In 2020, the lines gives 83.22.

16. Predict what year the life expectancy will be 92. When will the life expectancy be 105? (Round to the nearest year.)

    Answer: Solving the linear equation for 92, the year is 2053. For the life expectancy to be 105, the model predicts we will need to wait until 2102.

17. Does the residual plot (**Menu** > **Analyze** > **Residual** > **Show Residual Plot**) indicate that the linear regression is an accurate means of predicting life span? Explore other regressions to see if another model (such as an exponential or logistic model), more accurately fits the data. Which is best? Explain.

    Answer: The life expectancy data are not modeled best by exponential growth, but they are modeled by a logistic regression. Many real world population growth relationships are logistic. This means that the life expectancy will level off.