Lesson 12: Relationships Between Two Numerical Values

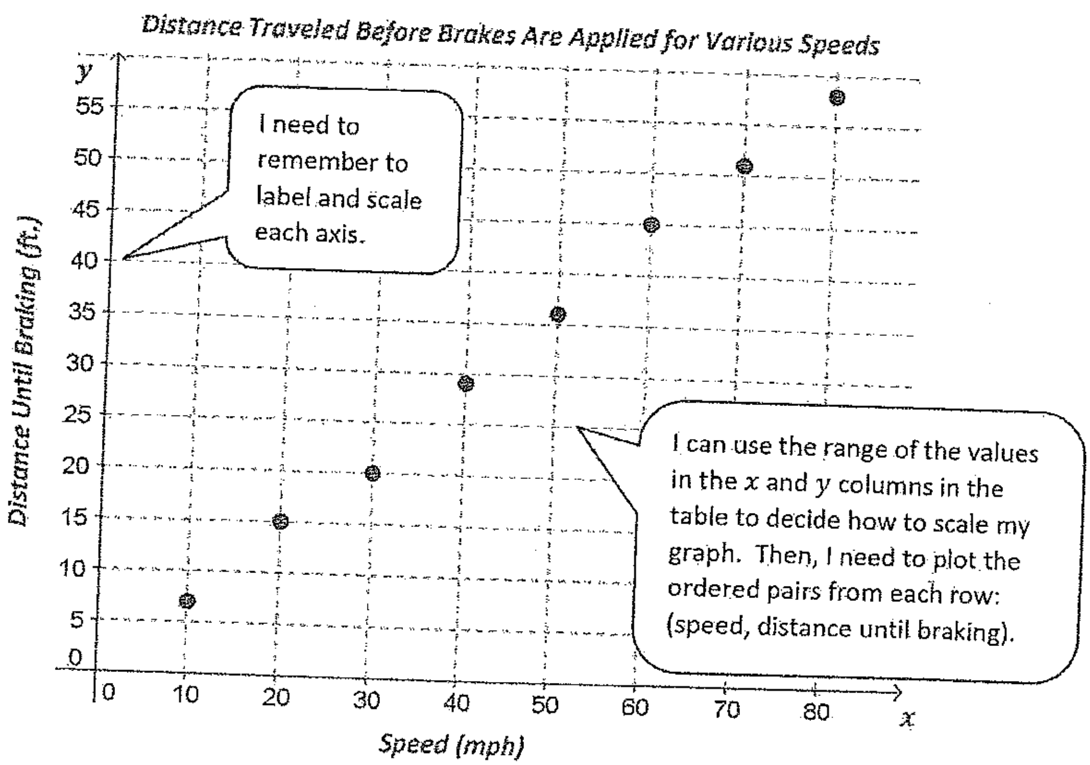
Construct a Scatter Plot and Analyze Relationships

The table below gives typical automobile braking road test results (distance traveled before brakes are applied and distance traveled until a complete stop after brakes are applied) for various speeds.

x (Speed in mph)	y (Distance Until Braking in ft.)	z (Distance Until Stopped in ft.)
10	7	5
20	15	17
30	20	37
40	29	65
50	36	105
60	45	150
70	51	205
80	58	265

(Data set from Core Math Tools, www.nctm.org)

1. Construct a scatter plot that displays the data where x represents speed (in mph) and y represents distance (in feet) traveled before braking.

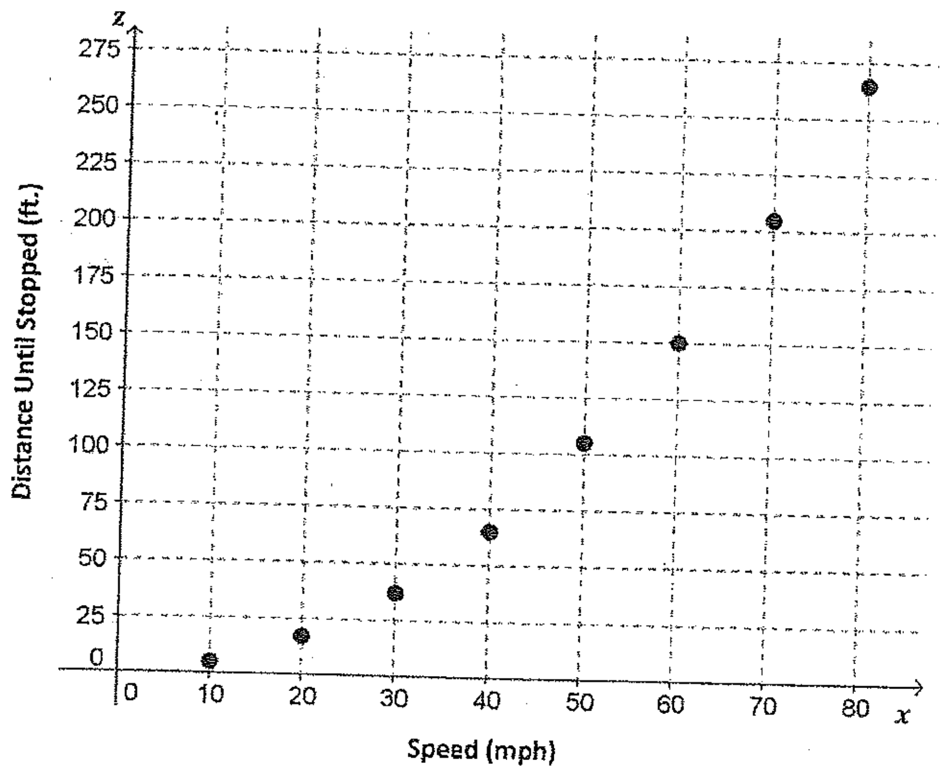


2. Based on the scatter plot, is there a relationship between the speed and the distance until braking? If so, how would you describe the relationship? Explain your reasoning.

There appears to be a relationship. As the speed increases, the distance until braking increases. The pattern suggests a linear model since the distance until braking increases at a nearly constant rate as the speed increases every 10 **mph**.

Consider the scatter plot where x represents the speed in mph and z represents the distance until stopped in feet shown below.

Distance Until Stopped After Brakes Are Applied for Various Speeds



3. Is there a relationship between the speed of an automobile and the distance until stopped after breaking or are the points scattered?

As the speed increases, the distance until stopped increases by larger and larger intervals. There is a pattern, so the points are not scattered.

4. Do you think there is a relationship between the speed and the distance until stopped? If so, does it look linear?

The relationship does not appear to be linear.

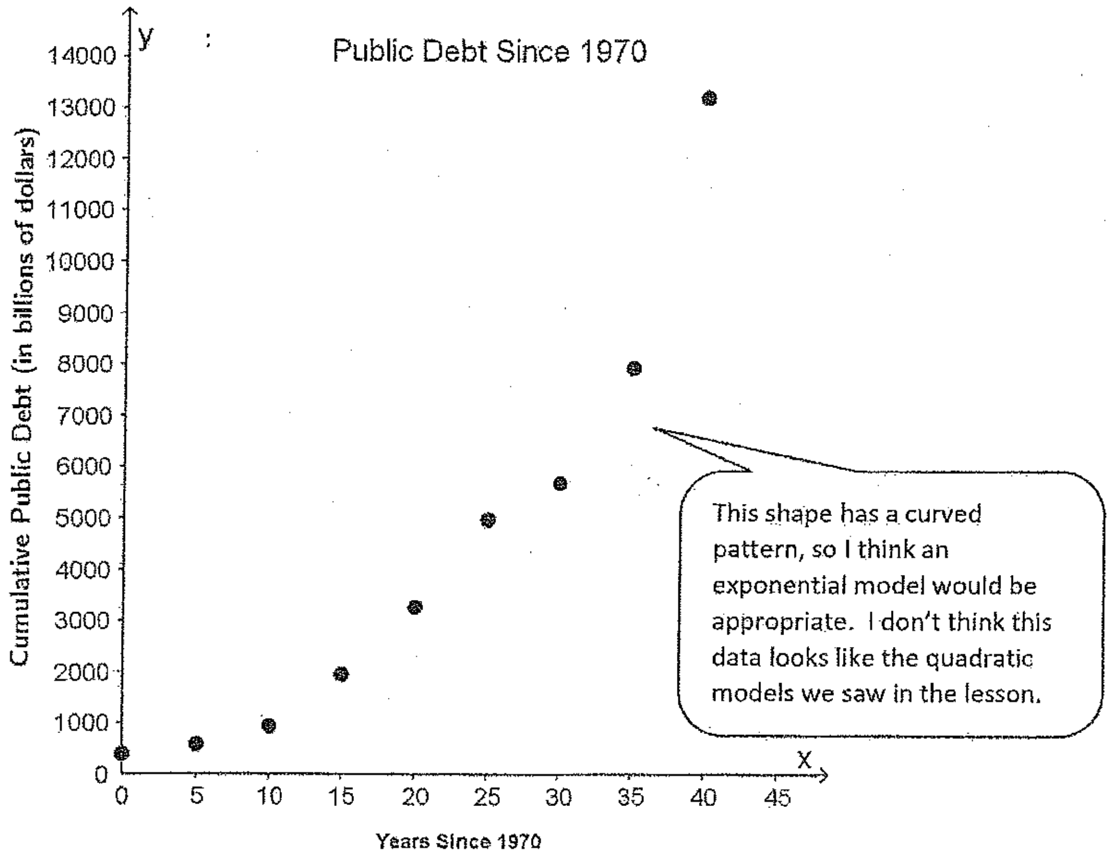
EUREKA MATH

Lesson 13: Relationships Between Two Numerical Values

Select a Model and Make Predictions

The scatter plot below shows the cumulative public debt (in billions of dollars) of the United States government at five-year intervals since 1970. (Source: U.S. Department of the Treasury, The Public Debt Online.)

Let x represent the years since 1970, and let y represent the cumulative public debt in billions of dollars.



1. What type of model (linear, quadratic, or exponential) would you use to describe the relationship between years since 1970 and the cumulative public debt in billions of dollars?

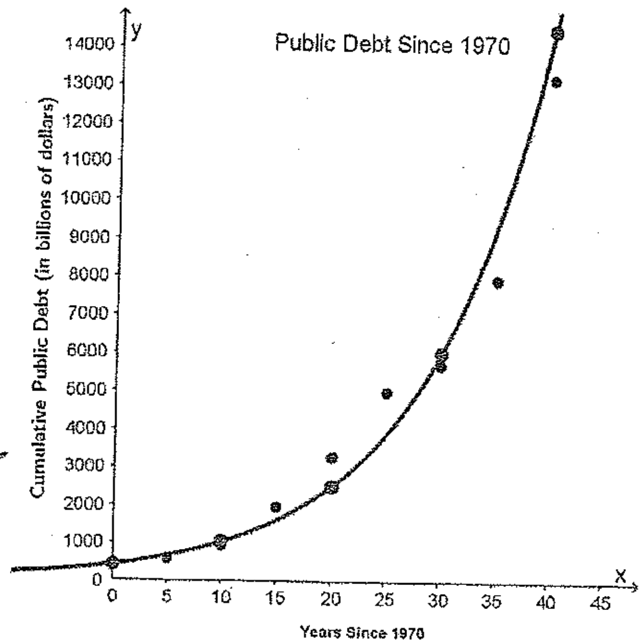
   An exponential model could be used to describe this relationship.

2. One model that could describe the relationship between the years since 1970 and the cumulative public debt is $y = 428.093(1.092)^x$. Use the exponential model to complete the table. Then sketch a graph of the exponential curve on the scatter plot above.

I need to substitute the years into the equation for x and use a calculator to evaluate and find y. I can round to the nearest whole number.

Years since 1970	Public Debt in Billions of $
0	429
10	1032
20	2489
30	6001
40	14,469

To graph the model, I can plot the ordered pairs represented in the table and connect them with a smooth curve.



3. Based on this model, how much cumulative public debt would you predict in the year 2015?

The model predicts cumulative public debt of 22,468 *billion dollars in 2015.*

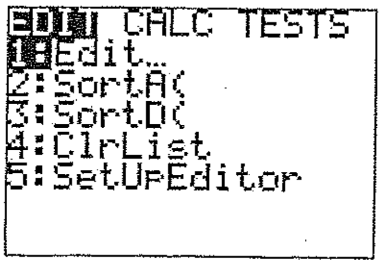
I know that 2015 is 45 years since 1970. I will substitute 45 into the model and find y.

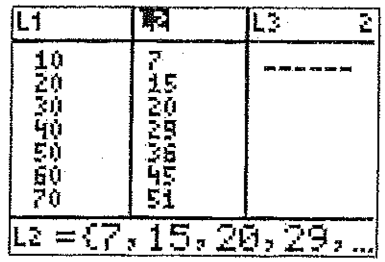
Lesson 14: Modeling Relationships with a Line

Lesson Notes

Finding the Regression Line (TI-84 Plus)

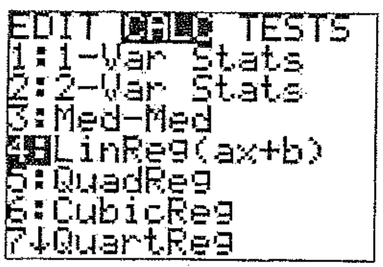
1. From your home screen, press STAT, and then from the STAT menu, select the EDIT option. (EDIT, ENTER)
2. Enter the x-values of the data set in L1, and enter the y-values of the data set in L2.

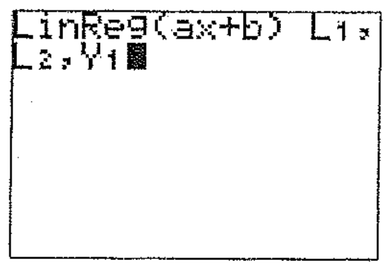




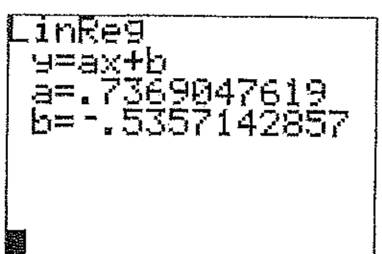
3. Select STAT. Move the cursor to the menu item CALC, and then move the cursor to option 4: LinReg($ax + b$) or option 8: LinReg($a + bx$). Press ENTER. (Note: Both options 4 and 8 are representations of a linear equation.)
4. With option 4 or option 8 on the screen, enter L1, L2, and Y1 as described in the following notes.

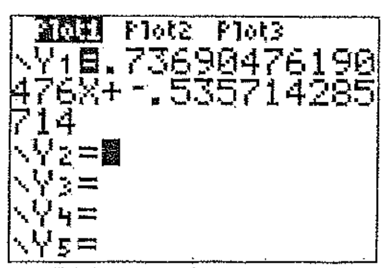
   LinReg($a + bx$) L1, L2, Y1 and select ENTER to see results.





To obtain Y1, go to VARS, and then move the cursor to Y-VARS and then to Functions (ENTER). You are now at the screen highlighting the y-variables. Move the cursor to Y1, and hit ENTER. Y1 is the least squares regression line and will be stored in Y1. To see the scatter plot, move the cursor to Plot1, and press ENTER.

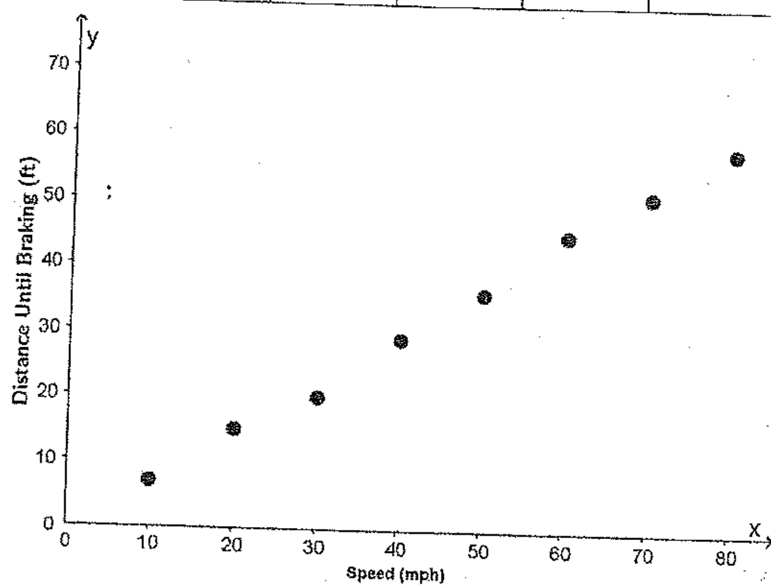




Find a Least Squares Regression Line

The table and scatter plot below give the typical distance an automobile will travel once a driver decides to hit the brakes before the driver actually engages the brakes. (Data set from Core Math Tools, www.nctm.org)

x is Speed (mph)	10	20	30	40	50	60	70	80
y is Distance Until Braking (ft.)	7	15	20	29	36	45	51	58



I can follow the steps on the previous page to create the least squares line on a graphing calculator.

1. Find the equation of the least squares line. (Round values to the nearest hundredth.)

   The least squares line is $y = 0.74x - 0.54$.

Desmos.com is a free online graphing calculator that can also be used to create the least squares line. There is a tutorial on their website that I can use to guide me through the steps.

I need to substitute 55 and 100 into the least squares line.

2. Predict the distance until braking 70 ft. for a car traveling 55 mph? What would you predict for a car traveling 100 mph?

   When $x = 55$, $y = 0.74(55) - 0.54 = 40.16$. *When* $x = 100$, $y = 0.74(100) - 0.54 = 73.46$. *The car travels approximately* 40 *feet before the brakes are applied when traveling* 55 mph *and approximately* 73 *feet before the brakes are applied when traveling* 100 mph.

3. Calculate the predicted value and the residual value for a car traveling 50 mph, and add it to the table below. Then calculate the sum of the squared residuals.

   $y = 0.74(50) - 0.54 = 36.46$

   Residual: $36 - 36.46 = -0.46$

   I need to subtract the predicted value from the actual value.

   Sum of squared residuals:

   $0.14^2 + 0.74^2 + (-1.66)^2 + (0.06)^2 + (-0.46)^2 + 1.14^2 + (-0.26)^2 + (-0.66)^2 = 5.3408$

Speed (mph)	Actual Distance Until Braking (ft.)	Predicted Distance Until Braking (ft.)	Residual
10	7	6.86	0.14
20	15	14.26	0.74
30	20	21.66	−1.66
40	29	29.06	−0.06
50	36	36.46	−0.46
60	45	43.86	1.14
70	51	51.26	−0.26
80	58	58.66	−0.66

I remember that slope is a rate of change. It is the coefficient of x in the least squares line.

4. Provide an interpretation of the slope of the least squares line.

   We would predict an additional 0.74 feet before braking for a speed increase of 1 mph.

5. Does it make sense to interpret the y-intercept of the least squares line in this context?

   No. The distance cannot be negative. The y-intercept is close to 0, which is the value that would make sense. When the speed is 0 mph, the car is not moving so the distance would also be 0 ft.

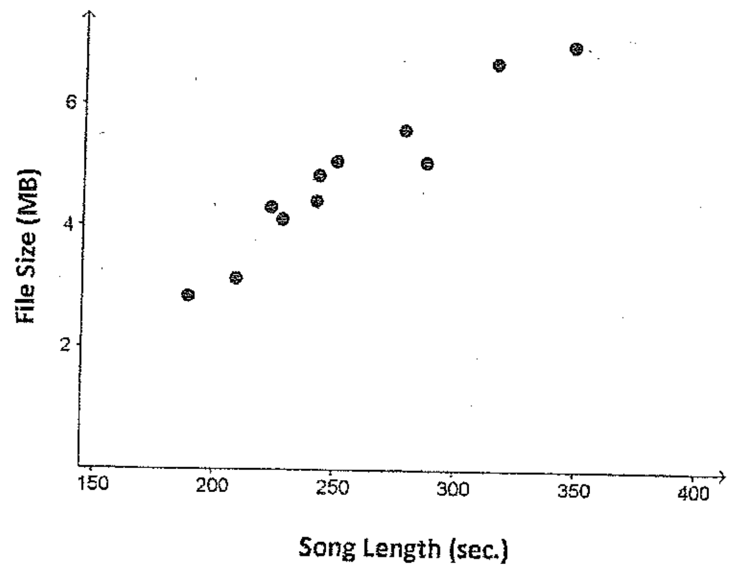
6. Would the sum of the residuals for the line $y = 0.9x - 1$ be greater than, about the same as, or less than the sum you computed in Problem 3?

   It would be greater because the least squares line has the smallest sum of the residuals of any linear model.

Lesson 15: Interpreting Residuals from a Line

The song length in seconds and file size in MB for several songs is shown in the table and scatter plot.

Song Length (sec.)	File Size (MB)
228	4.103
223	4.306
250	5.071
243	4.835
278	5.595
242	4.410
348	6.975
316	6.683
287	5.062
189	2.846
209	3.134



The equation for the least squares line is $y = 0.026x - 1.859$ where x is the time in seconds and y is the file size in megabytes (MB). (Data set from Core Math Tools, www.nctm.org)

1. Draw the least squares line on the graph.

   The line is drawn on the graph in the solution for Exercise 5.

   Let $x = 200$*; then* $y = 0.026(200) - 1.859 = 3.341$.

   Let $x = 300$*; then* $y = 0.026(300) - 1.859 = 5.941$.

   I need to determine the coordinates of two points on the line and then plot the points.

2. Interpret the slope of the least squares line.

   The slope is the coefficient of x.

   The slope is 0.026 MB *per second. For each additional second, the file size increases by* 0.026 MB.

3. What does the least squares line predict for the file size of a song that is 250 seconds long?

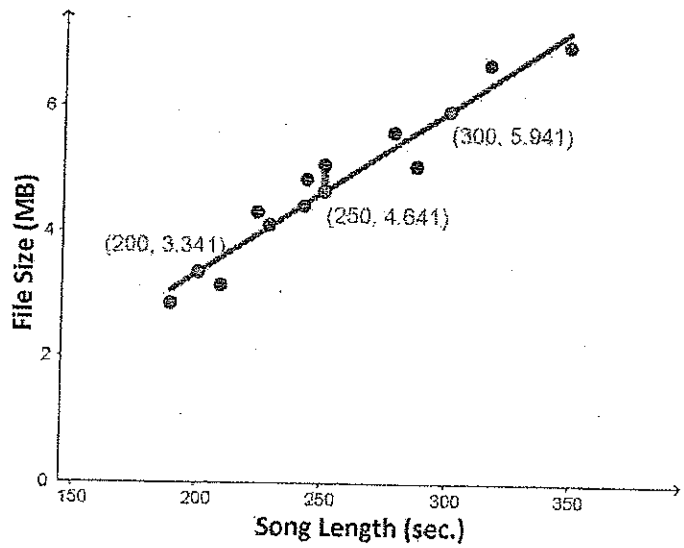
   When $x = 250$, $y = 0.026(250) - 1.859 = 4.641$. *The predicted size is* 4.641 MB.

4. What is the difference between the actual file size of a 250 second song and the predicted file size? (This is the residual.)

$5.071 - 4.641 = 0.43$

5. Show your answer to Exercise 4 as a vertical line between the point on the scatter plot and the least squares line.

I need to plot the point (250, 4.641) and connect that point to the point (250, 5.071). The length of the segment is 0.43.



6. Calculate all the residuals, and write them in the table below.

Song Length (sec.)	Actual File Size (MB)	Predicted File Size (MB)	Residual
228	4.103	4.069	0.034
223	4.306	3.939	0.367
250	5.071	4.641	0.430
243	4.835	4.459	0.376
278	5.595	5.369	0.226
242	4.410	4.433	−0.023
348	6.975	7.189	−0.214
316	6.683	6.357	0.326
287	5.062	5.603	−0.541
189	2.846	3.055	−0.209
209	3.134	3.575	−0.441

Using the graphing calculator, the residuals are stored in a list named RESID. I can insert this list where I enter other lists. Or, I can just calculate the residuals one at a time like I did in Exercises 3 and 4.

7. What does the least squares line predict for the file size of a 350 second song?

When $x = 350$*,* $y = 0.026(350) - 1.859 = 7.241$*. The predicted size is* 7.241 MB*.*

> I need to compare this residual to the others in the table. I can see it is quite a bit larger.

8. Would you be surprised if the actual file size of a 350 second song was 8 MB? Why or why not?

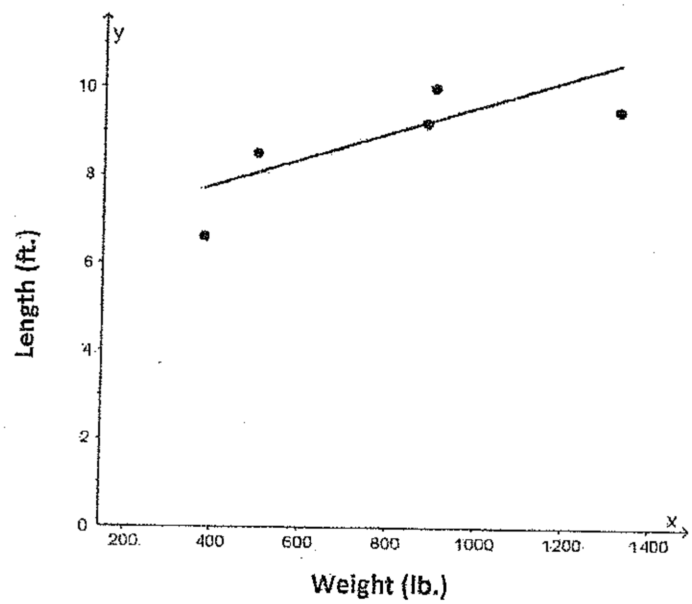
The residual would be $8 - 7.241 = 0.759$*. This number is larger than the other residuals, so a file this large would be surprising for a* 350 *second song.*

Lesson 16: More on Modeling Relationships with a Line

Create a Residual Plot

The lengths (in feet) and weights (in pounds) of five types of seals are shown in the table and scatter plot. (Source: "Grzimek's Encyclopedia, Mammals" V4. New York: McGraw-Hill, 1990, accessed via Core Math Tools, www.nctm.org)

Seal	Weight (in lb.)	Length (in ft.)
Crabeater	496	8.5
Harbor	375	6.6
Hooded	900	10.0
Monk	881	9.2
Weddell	1,323	9.5



I learned how to create the least squares line using a graphing calculator in Lesson 14.

The equation for the least squares line is $y = 0.003x + 6.58$, where x is the weight in pounds and y is the length in feet, and is included on the scatter plot.

1. Use your equation to find the predicted length of the hooded seal. What is the residual?

   When $x = 900$, $y = 0.003(900) + 6.58 = 9.280$. *The predicted length is* 9.280 ft.

   Residual = *Actual* − *Predicted* = $10.0 - 9.28 = 0.82$

2. Calculate the residuals for the other seals. Write the residuals in the table below.

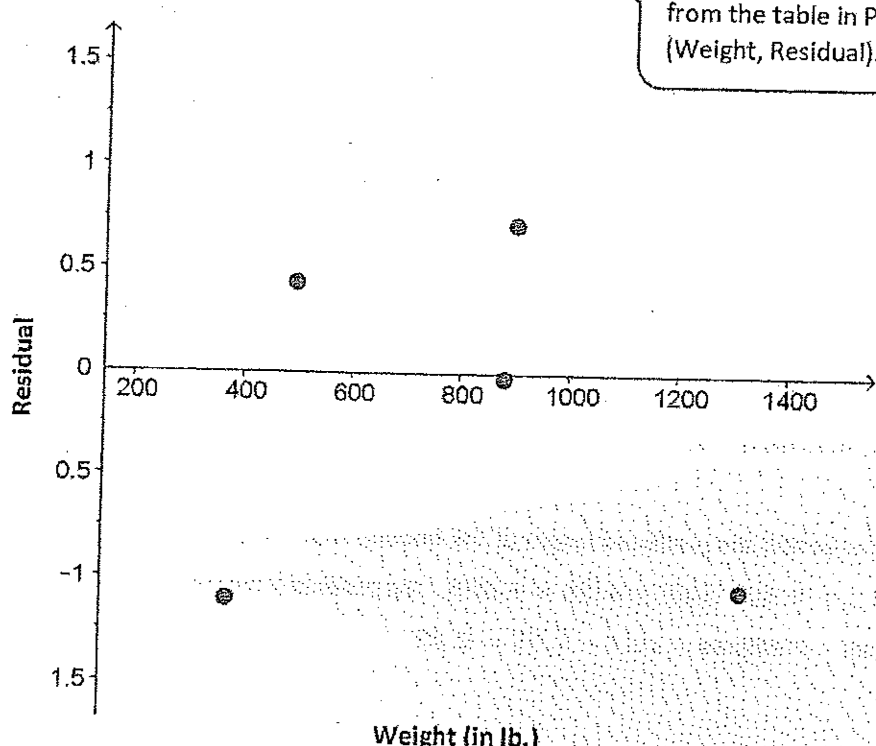
I can use the least squares line to get the predicted length.

Seal	Weight (in lb.)	Actual Length (in ft.)	Predicted Length (in ft.)	Residual
Crabeater	496	8.5	8.068	0.432
Harbor	375	6.6	7.705	−1.105
Hooded	900	10.0	9.280	0.720
Monk	881	9.2	9.223	−0.023
Weddell	1323	9.5	10.549	−1.049

3. Using the axes provided below, construct a residual plot for this data set.

I need to plot the ordered pairs from the table in Problem 2: (Weight, Residual).

The residuals are shown below.



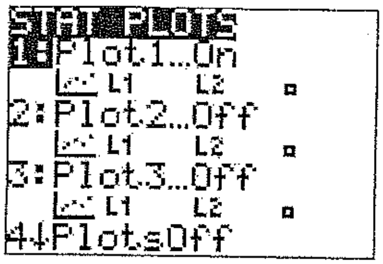
Lesson 17: Analyzing Residuals

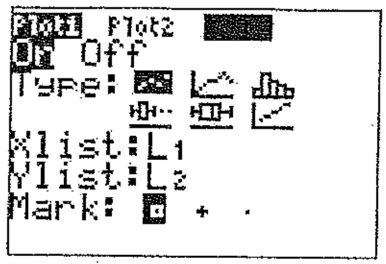
Lesson Notes

Students need access to a graphing calculator or other technology to complete this problem set. Directions for using a TI-84 Plus graphing calculator are listed below. Other types of calculators or tools may give slightly different values when calculating regression equations. Students can also use a spreadsheet program, free mathematics graphing software such as Geogebra, or a web-based application such as Desmos.com.

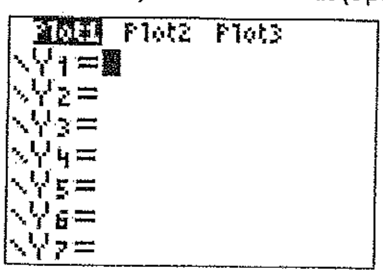
Construction of Scatter Plot:

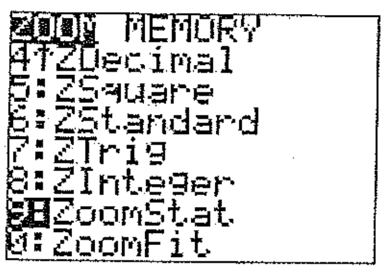
1. From the home screen press 2nd, STAT PLOT, and then select Plot1, and press ENTER.
2. Select *On*. Under *Type*, choose the first (scatter plot) icon, for Xlist enter L_1, for Ylist enter L_2, and under *Mark*, choose the first (square) symbol.

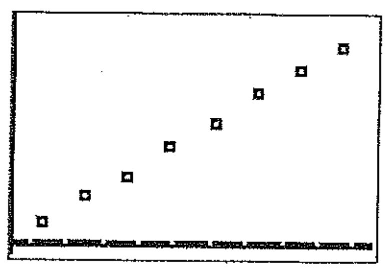




3. Press 2nd, QUIT to return to the home screen.
4. Press Y = and go to any unwanted graph equations and press CLEAR. Make sure that only Plot1 is selected (not Plot2 or Plot3).
5. Press Zoom, select ZoomStat (option 9), and press ENTER.







*these are general directions. The data is not the same as that given in the exercises on next pages.

Create a Scatter Plot and Least Squares Line

The table below shows the hippopotamus population sizes for various years. (Data accessed from Core Math Tools, www.nctm.org)

I will let the x represent the year and y represent the population.

Year	Hippopotamus Population
1970	2,815
1972	2,919
1975	2,342
1976	4,501
1977	5,147
1978	4,765
1979	5,151
1981	4,884
1982	6,293
1983	6,544

After I enter the data into two lists by pressing STAT and **Edit** and then typing in the data, I need to set up a scatter plot on my calculator in **STATPLOT** (2nd Y =).

To view the scatter plot on my calculator, I need to set up an appropriate viewing window by pressing **WINDOW**. The table values can be used to select appropriate **xmin, xmax, ymin,** and **ymax** values.

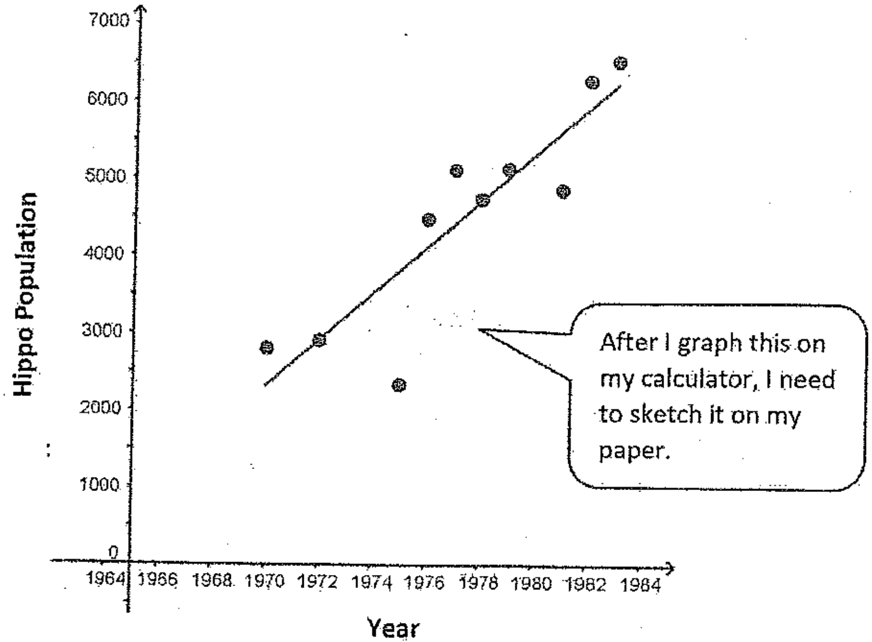
I learned the steps to create the least squares line in Lesson 14.

1. Use a calculator or computer to construct the scatter plot of this data set. Include the least squares line on your graph. Explain what the slope of the least squares line indicates about the hippopotamus population.

   The least squares equation is $y = 301.3x - 591231.4$, *where* x *is the year and* y *is the hippopotamus population for that year.*

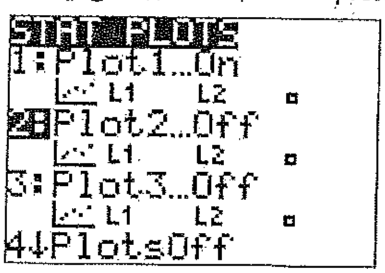
   The slope means that the population of hippos is increasing by about 301 *additional hippos each year.*

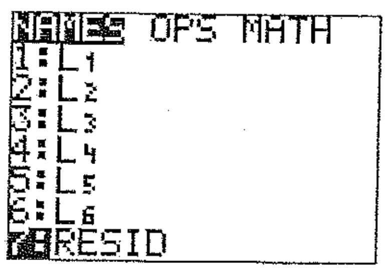
EUREKA MATH

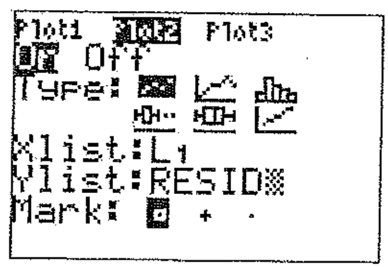


Construction of Residual Plot:

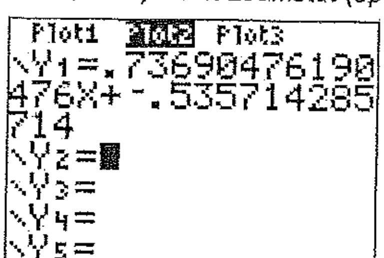
1. From the home screen, press 2nd, STAT PLOT, and then select Plot2, and press ENTER.
2. Select *On*. Under *Type*, choose the first (scatter plot) icon, for Xlist enter L_1, for Ylist enter RESID, and under *Mark* choose the first (square) symbol. (RESID is accessed by pressing 2nd, LIST, selecting NAMES, scrolling down to RESID and pressing ENTER.)

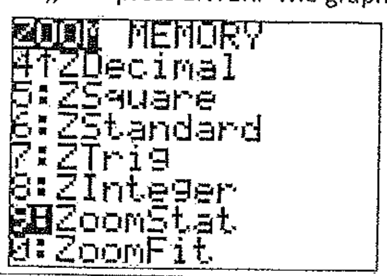


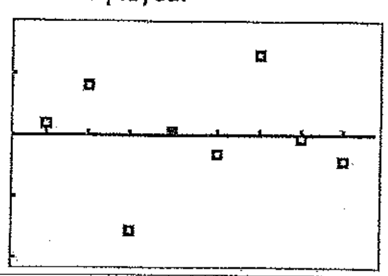




3. Press Y =. First, deselect the equation of the least squares line in Y_1 by going to the = sign for Y_1 and pressing ENTER. Then deselect Plot1, and make sure that Plot2 is selected.
4. Press Zoom, select ZoomStat (option 9), and press ENTER. The graph will be displayed.







Construct and Analyze a Residual Plot

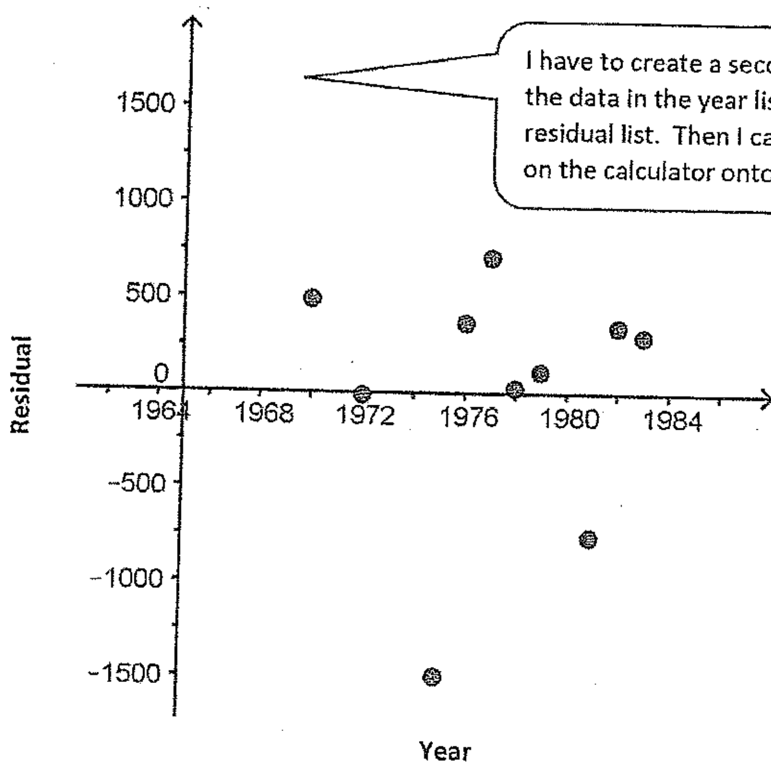
2. Use your calculator to construct a residual plot for this data set, and make a sketch on the axes given below. Does the scatter points on the residual plot indicate a linear relationship in the original data set? Explain your answer.

Year	Actual Population	Predicted Population	Residual
1970	2,815	2330	485
1972	2,919	2932	−13
1975	2,342	3836	−1494
1976	4,501	4137	364
1977	5,147	4439	708
1978	4,765	4740	25
1979	5,151	5041	110
1981	4,884	5644	−760
1982	6,293	5945	348
1983	6,544	6247	298

The graphing calculator stores residuals in a list named RESID. I will make a list for the predicted population and one for the residuals. I learned in Lessons 15 and 16 that:
residual = actual − predicted.

I got these values from my calculator and rounded them to the nearest whole number.

There is not a clear pattern in the data, so the residual plot would indicate a linear relationship.



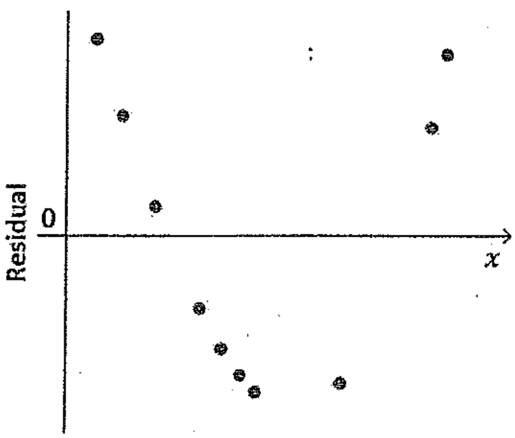
I have to create a second statplot using the data in the year list and data in the residual list. Then I can sketch what I see on the calculator onto my paper.

Lesson 18: Analyzing Residuals

Analyzing a Residual Plot

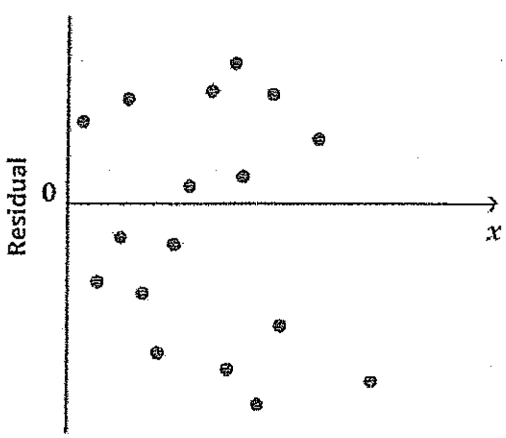
For each residual plot, what conclusion would you reach about the relationship between the variables in the original data set? Indicate whether the values would be better represented by a linear or nonlinear relationship.

1.



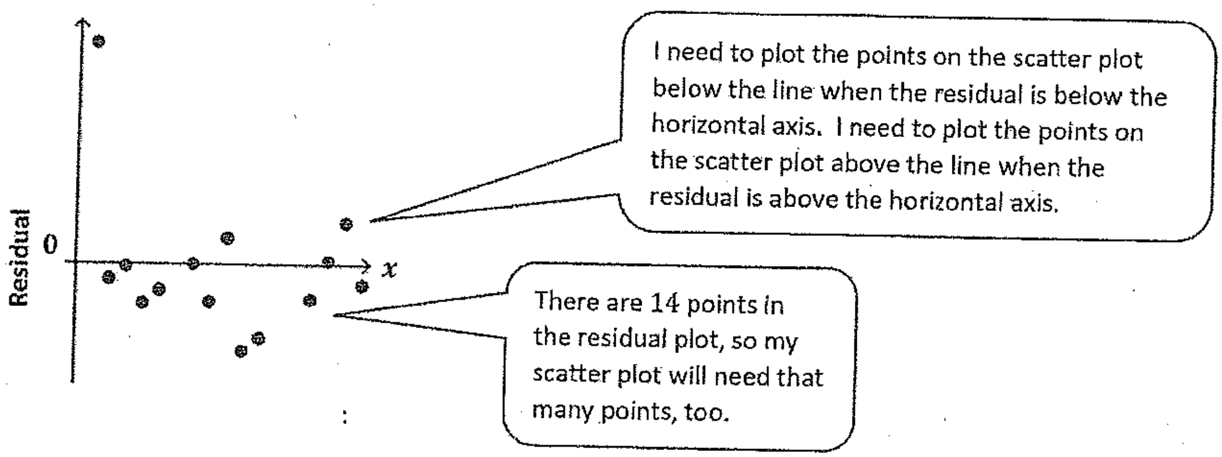
There is a pattern in the residuals for these data. The values would be better represented by a nonlinear relationship.

2.

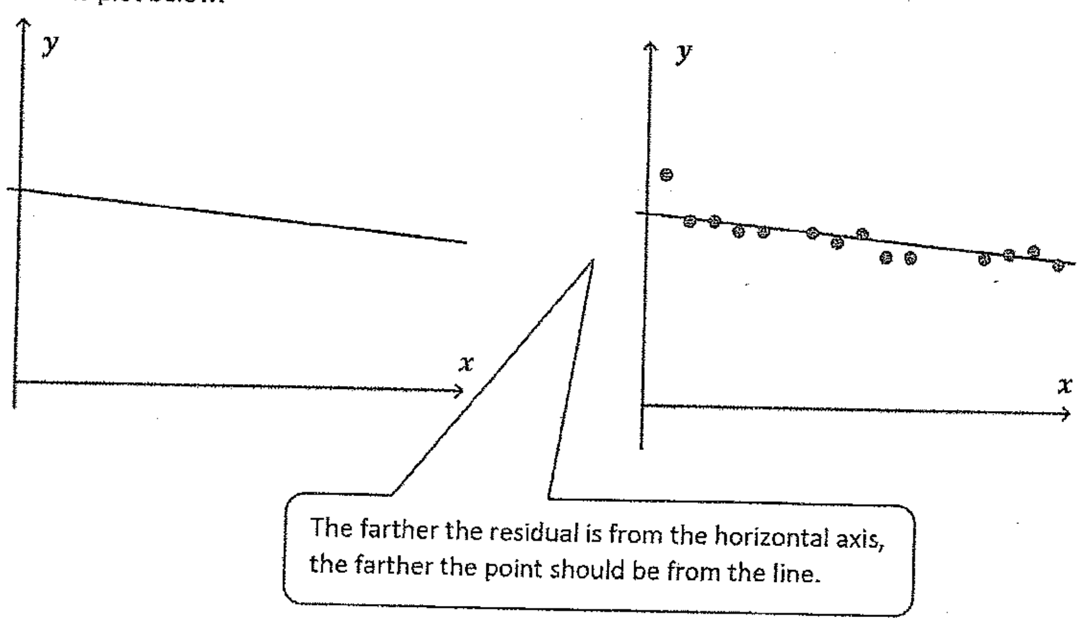


There is no pattern in the residuals. The values would be better represented by a linear relationship.

3. Suppose that after fitting a line, a data set produces the residual plot shown below.



An incomplete scatter plot of the original data set is shown below. The least squares line is shown, but the points are missing. Estimate the locations of the original points, and create an approximation of the scatter plot below.



EUREKA MATH

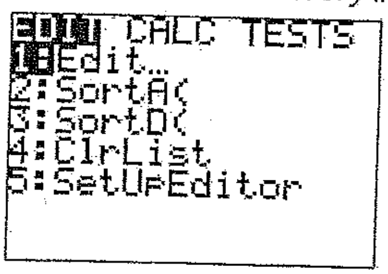
Lesson 19: Interpreting Correlation

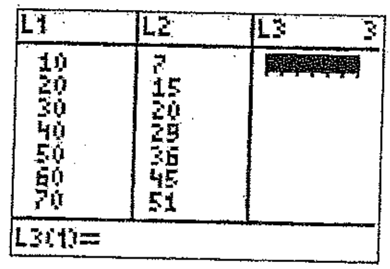
Lesson Notes

The steps below describe how to calculate a correlation coefficient using a TI-84 Plus graphing calculator. Different types of calculators or tools may return slightly different values in regression equations. If using a different graphing calculator, graphing software, or other graphing applications, students will need to consult the user guides for their technology.

Steps for Calculating the Correlation Coefficient Using a TI-84 Plus:

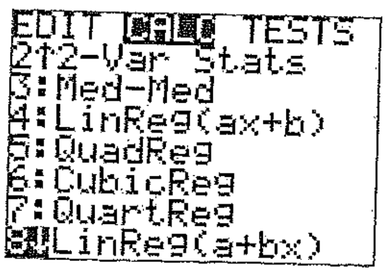
1. Determine which variable represents x and which variable represents y based on x- and y-variable designations.
2. From the home screen, select STAT and then Edit by pressing ENTER.
3. Enter the values of x in L1 and the values of y in L2. When complete, enter 2ND QUIT.

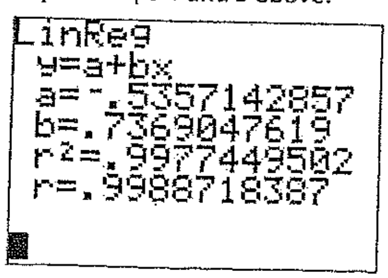


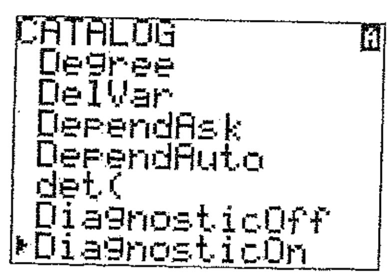


4. Select STAT. With the arrows, move the top cursor over to the option CALC, and move the cursor down to 8: LinReg($a + bx$) and then ENTER.
5. With LinReg($a + bx$) on the screen, enter L1, L2, Y1, and then ENTER. The value of r, the correlation coefficient, should appear on the screen. The least squares equation will be stored in Y1.

NOTE: If the r value does not appear, select 2ND CATALOG, move the cursor down to DiagnosticOn, and then ENTER. Press ENTER one more time. Repeat steps 4 and 5 above.

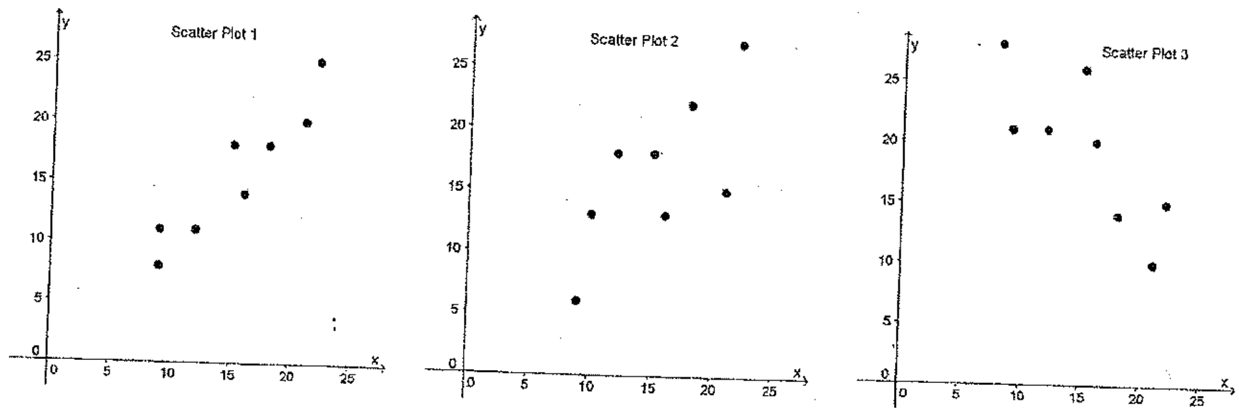






Identifying the Strength of a Linear Relationship

1. Which of these scatter plots below shows the strongest linear relationship? Which shows the weakest linear relationship?



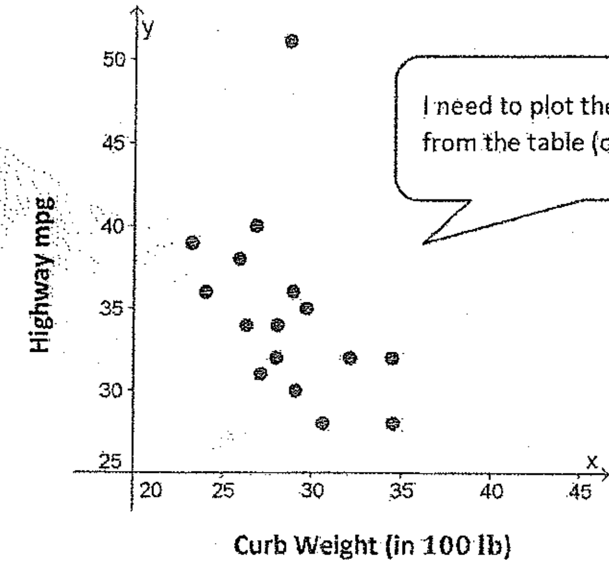
Scatter plot 1 has the strongest correlation. The weakest linear relationship is shown in scatter plot 2.

Calculating and Interpreting Correlation Coefficient

Edmunds.com published data on curb weight (in 100 lb) and highway miles per gallon for several compact car models.

Car Make and Model	Curb Weight (in 100 lb)	Highway mpg
Audi A4	34.50	32
Chevrolet Cobalt	32.16	32
Ford Focus	26.36	34
Honda Civic	26.90	40
Honda Civic Hybrid	28.75	51
Hyundai Accent	24.03	36
Kia Spectra	29.72	35
Mazda3	28.11	34
Mercedes-Benz C280	34.60	28
Nissan Sentra	28.97	36
Saturn ION	28.05	32
Subaru Impreza	30.67	28
Suzuki Aerio	27.16	31
Toyota Corolla	25.95	38
Toyota Yaris	23.26	39
VW Rabbit	29.11	30

2. Construct a scatter plot of these data using the grid provided.



I need to plot the ordered pairs from the table (curb weight, mpg).

After entering the data into lists, I calculate the least squares line. The correlation coefficient is given by the r value shown on my calculator screen.

3. Calculate the value of the correlation coefficient between curb weight and highway miles per gallon and interpret this value. Round to the nearest hundredth.

The correlation coefficient is $r = -0.43$. This indicates a moderately negative relationship between weight in hundreds of pounds and highway mpg.

The least squares line would have a negative slope. I know that means the correlation coefficient will be negative as well.

4. Does it surprise you that the correlation coefficient is negative? Explain why or why not?

No. Heavier cars should need more gas to move them forward provided other variables are equal such as tire quality, engine efficiency and size, type of fuel, etc.

5. Is it reasonable to conclude heavier cars have a higher highway miles per gallon?

No. It appears that as the cars get heavier, the highway miles per gallon decreases.

6. Is it reasonable to conclude that increasing the weight of a car decreases the highway miles per gallon?

No. Just because there is a correlation between these two variables does not mean there is a cause-and-effect relationship between the two.

I need to remember that correlation does not imply that one variable causes the changes in the other.