3.1

Is It Getting Hot in Here?

Modeling Data Using Linear Regression

LEARNING GOALS

In this lesson, you will:

- Create a graph of data points on a graphing calculator.
- Determine a linear regression equation using a graphing calculator.
- Recognize the accuracy of a line of best fit using the correlation coefficient.
- Make predictions about data using a linear regression equation.

KEY TERMS

- linear regression
- line of best fit
- linear regression equation
- significant digits
- correlation coefficient

PROBLEM 1 What's Your Prediction?

The table shown lists the average global temperature for each decade from 1880 to 2009.

Decade Number	Decade	Average Temperature (°F)
0	1880–1889	56.876
1	1890–1899	56.642
2	1900–1909	56.732
3	1910–1919	56.822
4	1920–1929	57.038
5	1930–1939	57.236

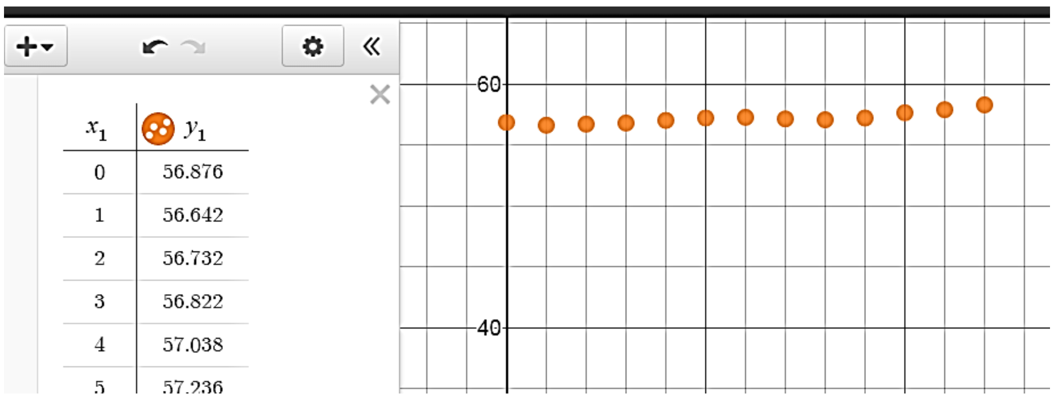
1. Identify the independent and dependent quantities and their units of measure.
   Independent Quantity (IQ): time — Units: decades (10 years)
   Dependent Quantity (DQ): temperature — Units: degrees (Fahrenheit)

2. Do the data represent a function? Why or why not? If so, describe the function.
   Yes, each decade number (#) is paired with exactly 1 average global temperature.
   The avg global temperature (DQ) is a function of the decade # (IQ).

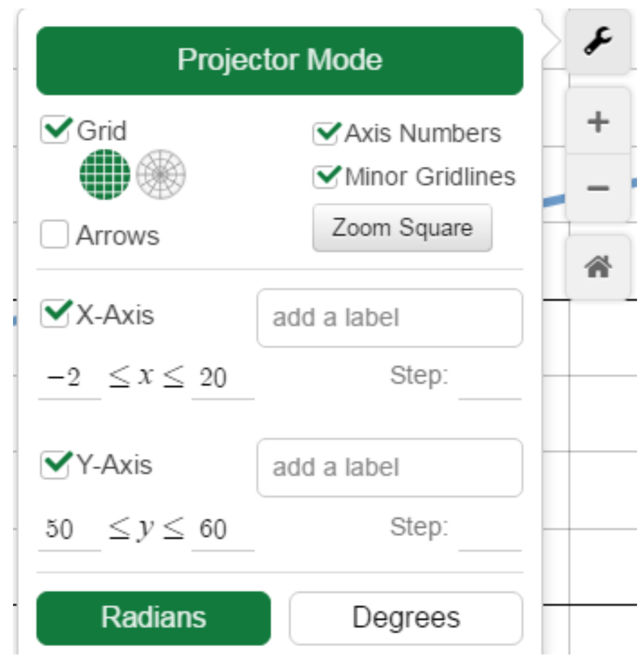
You can represent the data using a graphing calculator. In order to enter the data in your calculator, you must represent each decade as a single value.

3. Follow the steps provided to graph the relationship between time and temperature on a graphing calculator.

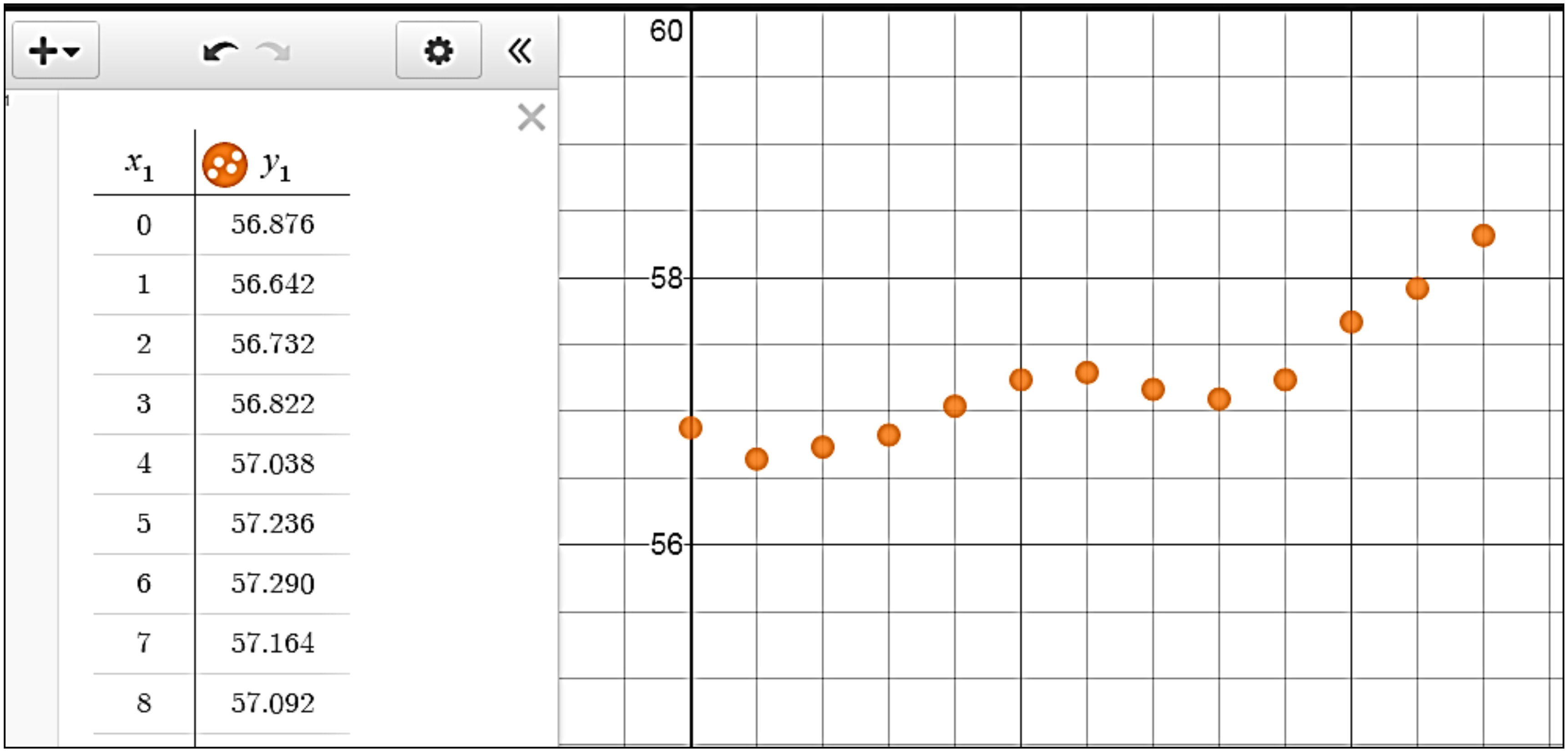
1. Go to www.desmos.com
2. Click “Start Graphing >”
3. Go to www.dbakeralgebra1.weebly.com - Course Content/Chapter 3 and “copy” the data table.
4. Go to Desmos and “paste” the data table in the upper left corner. You should see something like this…



5. Click on the wrench icon in the upper right corner and change your x-axis and y-axis to match the screen print to the right:



Your graph should now look like this:



4. Why do you think the first decade is numbered 0?

It's the *starting point* for the function, and it's the *y-intercept* for the graph.

5. Between which consecutive decades was there a decrease in average global temperature?

0 and 1, 6 and 7, & 7 and 8

6. What is the range of the data set?

Calculate the difference between the highest and lowest temperatures.
The range is 58.316° - 56.642° = 1.674°.

7. Is it possible to predict the approximate average global temperature for 2070 – 2079 using the graph? Explain your reasoning.

No, there are dips in the data which make it impossible to determine if the temperature will continue to increase through 2079.

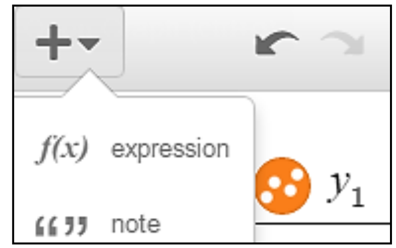
8. Would it make sense to draw a smooth curve connecting the points in the plot? Why or why not?

No, the data is discrete. There is only one average global temperature for each decade.

PROBLEM 2 Does That Seem to Fit?

Scientists often use a *linear regression* to model data in order to make predictions. A **linear regression** models the relationship between two variables in a data set by producing a *line of best fit*. A **line of best fit** is the line that best approximates the linear relationship between two variables in a data set. The equation that describes the line of best fit is called the **linear regression equation**.

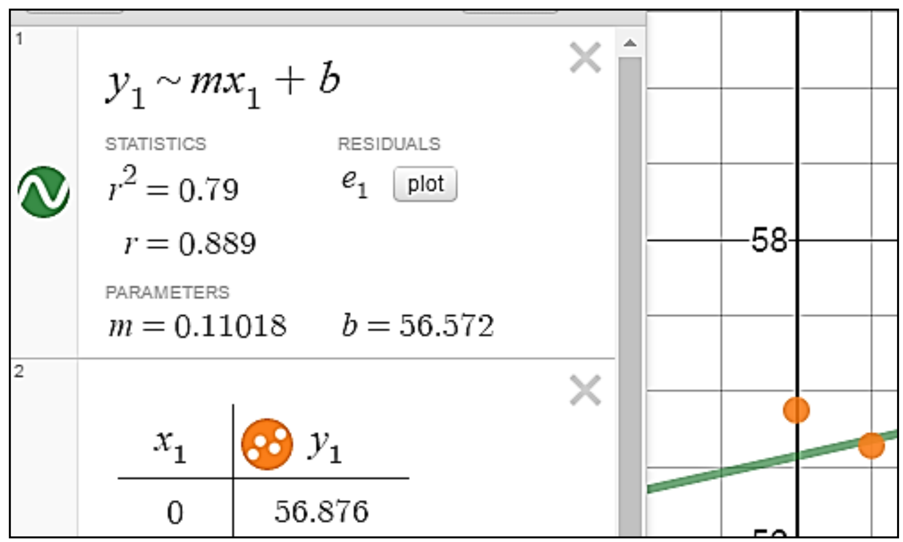
You can use a graphing calculator to determine a linear regression equation and then draw a line of best fit for the average global temperature data.

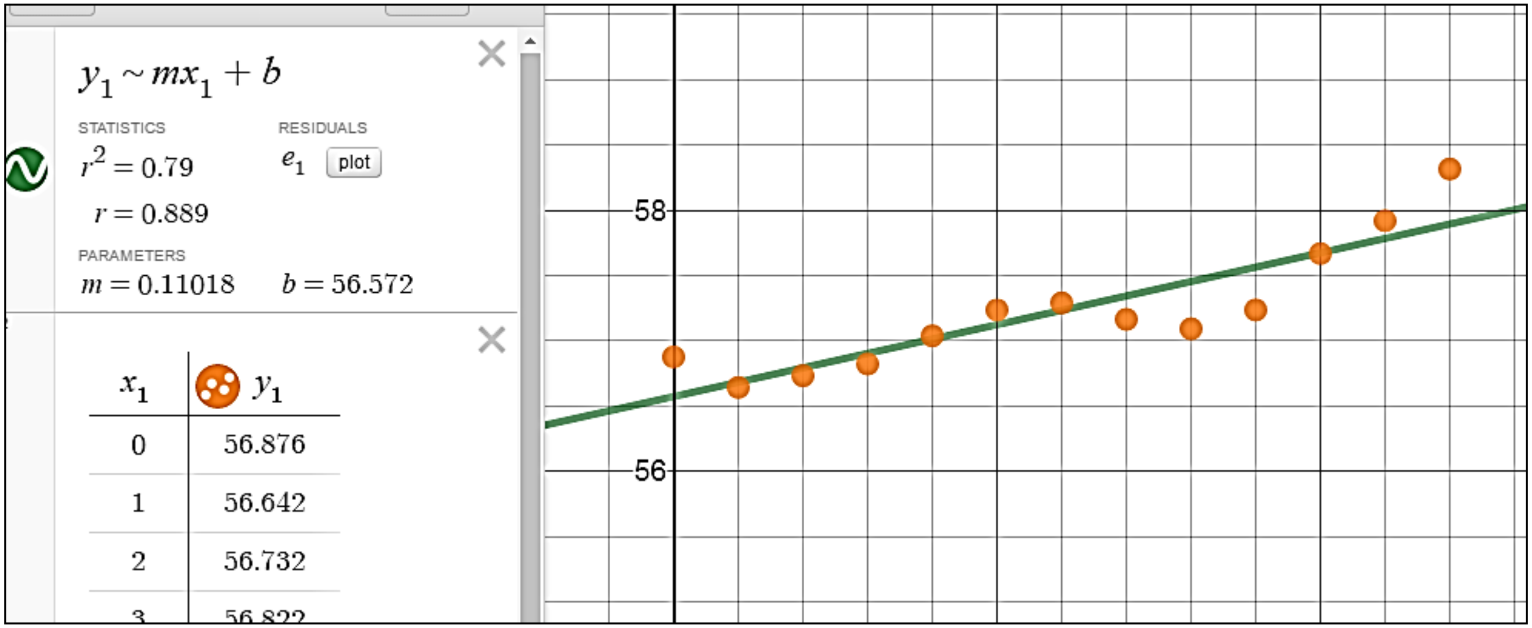


1. Go to the +▾ in the upper left corner and choose "*f(x)* expression" from the drop down list.

2. To add the expression: $y_1 \sim mx_1 + b$:
   Type: *y*1.
   Click "functions", "stats", then "~".
   Type: *mx*1 + *b*.

   You should see something that looks like the screen print to the right:





1. Determine the linear regression equation for the average global temperature data.

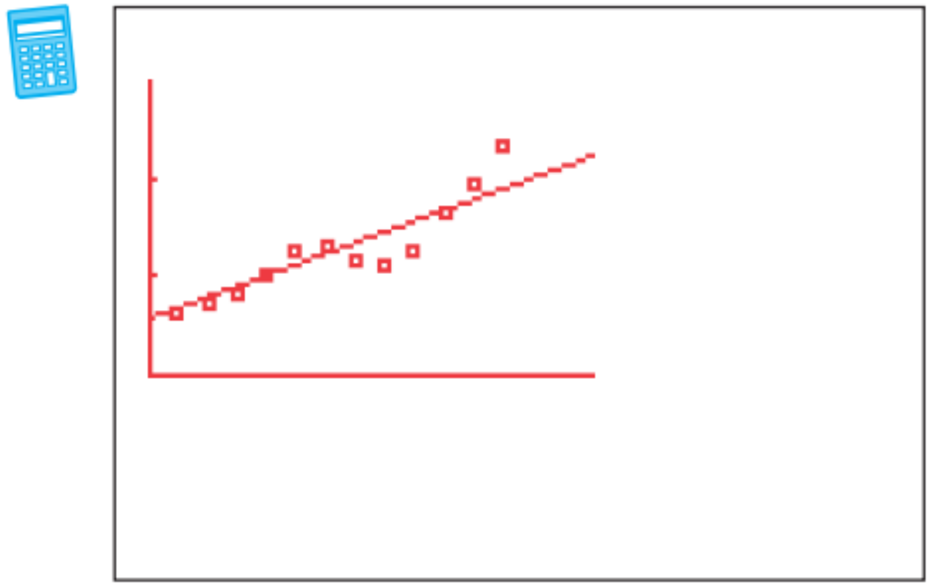
   $y = mx + b \longrightarrow y = 0.11018x + 56.572$

2. Rewrite the linear regression equation as a function. This time, round the slope and y-intercept to the appropriate place. Explain your reasoning.

   $f(x) = 0.110x + 56.572$

   The equation is rounded to 1000^{th} place because the data is rounded to the 1000^{th} place, but we will typically round to hundredths. $f(x) = 0.11x + 56.57$

3. Sketch the data points and the line of best fit that you see on the calculator.



a. Do the data show a positive correlation or a negative correlation? How can you tell?
Positive. The line of best fit slopes upward as it goes from left to right.

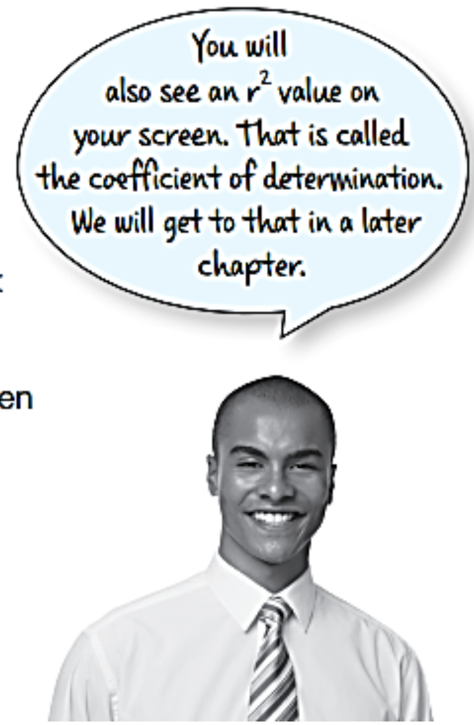
b. Do you think this line fits the data well? Explain your reasoning. Answers may vary.
Yes, because the line is close to most of the data points and it is going uphill.

The variable r on your linear regression screen is used to represent the *correlation coefficient*. The **correlation coefficient** indicates how closely the data points form a straight line.

If the data show a positive correlation, then the value of r is between 0 and 1. The closer the data are to forming a straight line, the closer the value of r is to 1.

If the data show a negative correlation, the value of r is between 0 and -1. The closer the data are to forming a straight line, the closer the r-value is to -1.

If there is no linear relationship in the data, the value of r is 0.



4. What is the correlation coefficient, or r-value, for your line of best fit? Interpret the meaning of the r-value.

$r = 0.889$
The r-value is positive so the data shows a positive correlation.
Since the r-value is close to 1, the line is a good fit.