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

Lines of regression are used to make predictions like who will be the next president or how global warming will affect us in the future.

PROBLEM 1 What's Your Prediction?

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

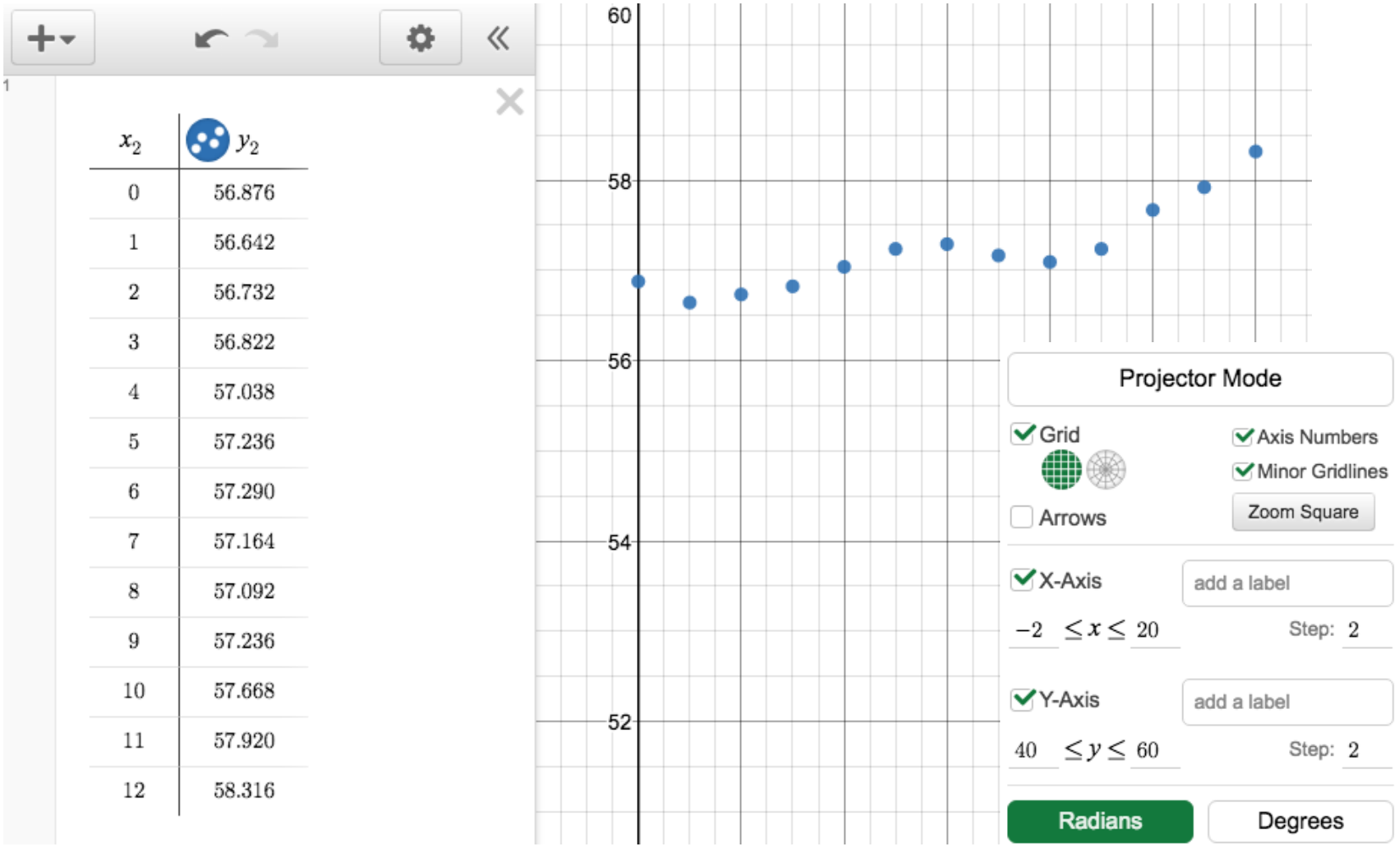
Decade Number	Decade	Average Temperature (°F)
0	1880–1889	56.876
1	1890–1899	56.642
2	1900–1909	56.732
3	1910–1919	56.822
4	1920–1929	57.038
5	1930–1939	57.236

1. Identify the independent and dependent quantities and their units of measure.
   IQ: Time (in decades, 10 years)
   DQ: Temperature (in degrees Fahrenheit)

2. Do the data represent a function? Why or why not? If so, describe the function.
   Does each decade map to 1 and only 1 temperature?
   Yes, each decade is paired with exactly 1 average global temperature.
   The temperature is a function of the decade.

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

3. We can use the graphing calculator, www.desmos.com, to enter the data table and create a scatter plot like the one below:



x_2	y_2
0	56.876
1	56.642
2	56.732
3	56.822
4	57.038
5	57.236
6	57.290
7	57.164
8	57.092
9	57.236
10	57.668
11	57.920
12	58.316

4. Why do you think the first decade is numbered 0?
   It's the *starting point* for the data, and it's the *y-intercept* for the graph.

5. Between which **consecutive** decades was there a **decrease** in average global temperature?
   Between decades 0 and 1, 6 and 7, & 7 and 8

6. What is the ***range*** of the data set?
   Hint: find the difference between the highest and lowest temperatures.
   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 making it impossible to predict if it will continue to increase through 2079.

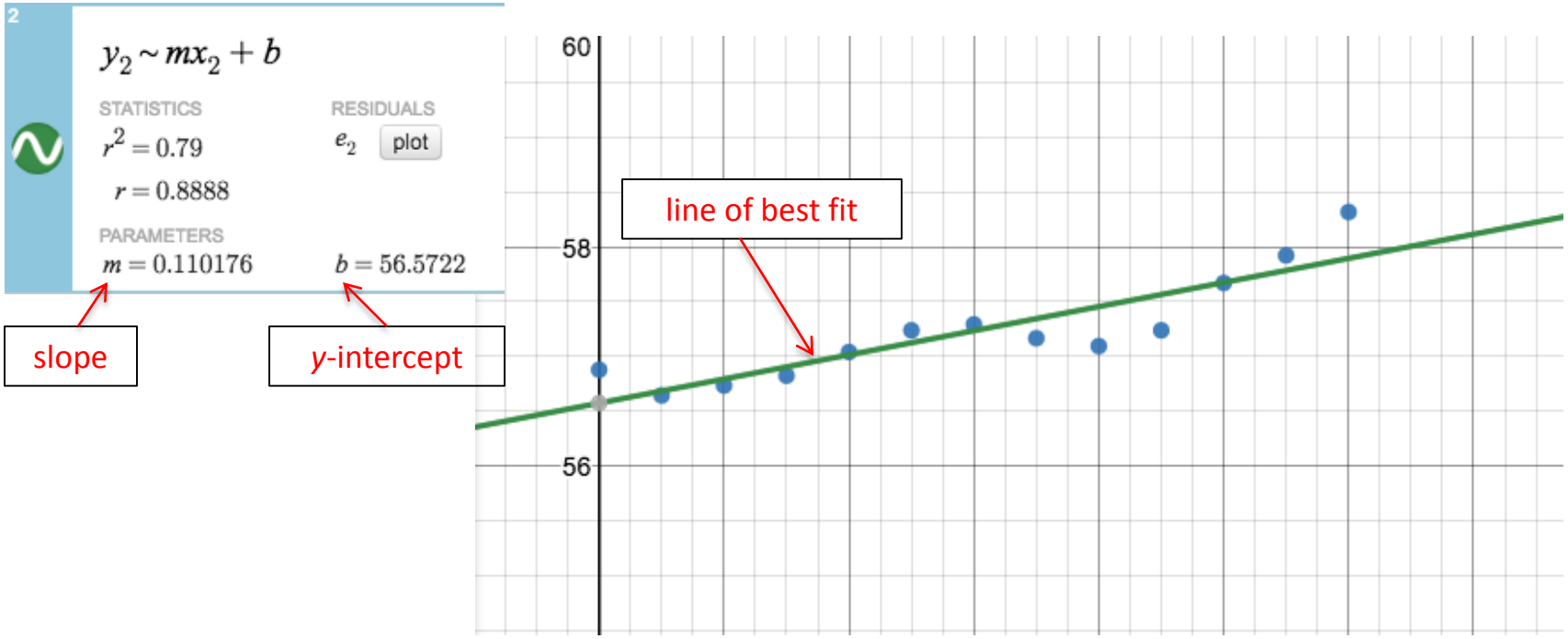
8. SKIP

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.

We can also use www.desmos.com to draw a line of best.



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

   m = 0.110176

   b = 56.5722

   $y = mx + b \longrightarrow y = 0.110176x + 56.5722$ or $f(x) = 0.110176x + 56.5722$

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

   $f(x) = 0.110x + 56.572$

   Round to the 1000th place since the data in the table is rounded to the 1000th place. *Typically, we round to the 100th place.*

x_2	y_2
0	56.876
1	56.642
2	56.732
3	56.822

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

   You do not need to sketch the graph since you already have a copy.

   a. Do the data show a positive correlation or a negative correlation? How can you tell?

      Positive. The line of best fit slopes upward.

   b. Do you think this line fits the data well? Explain your reasoning. Answers may vary.

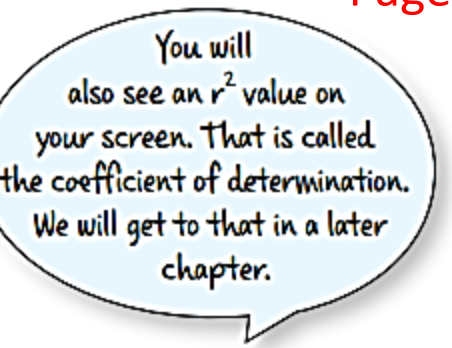
      Yes. The line is close to most of the data points.

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 = 0.8888
r is positive so the data has a positive correlation.
Since the *r*-value is close to 1, the line is a good fit.