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

You may have heard a lot of conversation recently about global warming. But did you know that just 40 years ago, some scientists thought the Earth was actually going through a time of global cooling?

In the 1970s, scientists became aware that the averages of global temperatures had decreased since the 1940s. Evidence shared at a climate conference in 1965 triggered thought that changes in sunlight might cause an ice age which would begin within a few thousand years. At that time, much less was known about world climate and a few severe winters in 1972 and 1973 had the public believing this theory.

These thoughts and theories were all based on data which was then used to predict trends. Do you think this is a reliable method? Why might scientists use this method, even if it is not always correct?

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
6	1940–1949	57.290
7	1950–1959	57.164
8	1960–1969	57.092
9	1970–1979	57.236
10	1980–1989	57.668
11	1990–1999	57.920
12	2000–2009	58.316

1. Identify the independent and dependent quantities and their units of measure.

2. Do the data represent a function? Why or why not? If so, describe the function.

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.

You can use a graphing calculator to represent a data set.

Step 1: Press **STAT** and then press **ENTER** to select **1:Edit**. In the **L1** column, enter the independent quantity values by typing each value followed by **ENTER**.

Step 2: Use the right arrow key to move to the **L2** column. Enter the dependent quantity values.

Step 3: Press **2ND** and **STAT PLOT**. Press **4** to turn off any plots. Press **ENTER**. Then press **2ND** and **STAT PLOT** again. Press **ENTER** to access the information about Plot 1. The cursor should be on the word **On**. Press **ENTER** to turn on Plot 1.

If there is already data in your L1 list, highlight the heading L1, Press CLEAR, then Press ENTER to delete it.

Remember to use the decade number for the independent quantity and not the actual decade years.

Step 4: Use the arrow keys to move down to **Xlist**. Press **2ND L1** to set your L1 values as your *x*-values. Scroll to **Ylist** and Press **2ND L2** to set your L2 values as your y-values.

Step 5: Press **WINDOW** to set the bounds of your graph. Press **GRAPH** to create a graph of the data.

Step 6: Use the **TRACE** feature and the left and right arrow keys to move between the points on the plot.

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

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

6. What is the range of the data set?

3

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

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

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.

You can use a graphing calculator to determine the linear regression equation of a data set.

Step 1: **Diagnostics** must be turned on so that all needed data is displayed. Press **2nd CATALOG** to display the catalog. Scroll to **DiagnosticOn** and press **ENTER**. Then press **ENTER** again. The calculator should display the word **Done**.

Step 2: Press **STAT** and use the right arrow key to show the **CALC** menu. Type **4** to choose **LinReg(ax+b)** and press **ENTER**.

Step 3: Make sure **L1** is listed next to **Xlist** and **L2** is listed next to **Ylist**. Scroll down to **Calculate** and press **ENTER**.

The calculator should show $y = ax + b$ as well as four values labeled **a**, **b**, **r²**, and **r**.

You will need to substitute the values of *a* and *b* into the equation to write the linear regression equation.

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

Your graphing calculator may provide numbers rounded to several decimal places for the values used in the linear regression equation. However, are all those decimal places necessary? Using *significant digits* can help you make more sense of the values. **Significant digits** are digits that carry meaning contributing to a number's precision.

In this problem situation, the temperature measurements are accurate to the thousandths place, so each temperature measurement has 3 significant decimal digits. When constructing equations and expressions based on data, it is appropriate to use numbers that have the same number of significant decimal digits.

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

Let's graph this linear regression equation to see if it is an appropriate fit for the data.

Step 4: Press **Y=**. Enter the linear regression equation next to **Y_1=**. Then press **GRAPH** to see the line of best fit.

The calculator will automatically copy the linear regression equation to Y1 if you enter LinReg(*ax* + *b*) **Y1**. Repeat Step 2 to enter LinReg(*ax* + *b*), then press **VARS** and use the right arrow keys to show **Y-VARS**. Press **FUNCTION** and select **Y1**. When you press **ENTER** the equation will appear in **Y1**.

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?

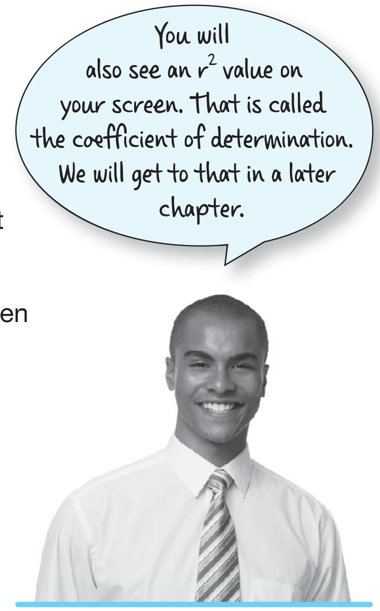
b. Do you think this line fits the data well? Explain your reasoning.

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.

3

5. Compare your sketch with your classmates' sketches. Locate a classmate who used bounds different from your bounds. How does changing the **Xmin** and **Xmax** change the look of the graph? How does changing the **Ymin** and **Ymax** change the look?

PROBLEM 3 Analyzing Linear Regression

1. For each expression from your linear regression equation, write an appropriate unit of measure and describe the contextual meaning. Then, choose a term from the word box to describe the mathematical meaning of each part.

output value	rate of change	input value	*y*-intercept

		What It Means	
Expression	Unit	Contextual Meaning	Mathematical Meaning
$f(x)$			
0.11			
x			
56.572			

Use your linear regression equation to answer each question.

2. About how much was the average global temperature changing each decade from 1880 to 2009 according to the data? Explain how you know.

3

3. Compare the *y*-intercept from the table with the *y*-intercept from the linear regression equation. What do you notice? Does this make sense in terms of the problem situation? Why or why not?

4. Use your equation to predict the average global temperature for the years 2070–2079. Show your work and explain your reasoning.

3

5. Predict the first decade with an average global temperature of at least 60°F.

Be prepared to share your solutions and methods.

3