

# Build a function that models a relationship between two quantities. MCC9-12.F.BF.1, 1a, 1b, 2

## Standards

**MCC9-12.F.BF.1** Write a function that describes a relationship between two quantities. \*

**MCC9-12.F.BF.1a** Determine an explicit expression, a recursive process, or steps for calculation from a context .

**MCC9-12.F.BF.1b** Combine standard function types using arithmetic operations. *For example, build a function that models the temperature of a cooling body by adding a constant function to a decaying exponential, and relate these [functions](#) to the model.*

**MCC9-12.F.BF.2** Write arithmetic and geometric [sequences](#) both recursively and with an explicit formula, use them to model situations, and translate between the two forms. \*

Suppose that each day an ice cream truck recorded the high temperature outside and the number of ice cream treats sold. They then entered the data points into a graphing calculator. Do you think that they could find the equation of the line that best fit the data? If so, how would they do it? Could they make predictions based on the equation? After completing this Concept, you'll be able work on problems like this one by entering data points into a graphing calculator and finding the line of best fit.

---

## Guidance

The real-world situations you have been studying so far form [linear equations](#). However, most data in life is messy and does not fit a line in [slope](#)-intercept form with 100% accuracy. Because of this tendency, people spend their entire career attempting to fit lines to data. The equations that are created to fit the data are used to make predictions, as you will see in the next Concept.

This Concept focuses on graphing **scatter plots** and using a scatter plot to find a linear equation that will best fit the data.

A **scatter plot** is a plot of all the ordered pairs in a table. This means that a scatter plot is a relation, and not necessarily a function. Also, the scatter plot is **discrete** , as it is a set of distinct points. Even when we expect the relationship we are analyzing to be linear, we should not expect that all the points would fit perfectly on a straight line. Rather, the points will be “scattered” about a straight line. There are many reasons why

the data does not fall perfectly on a line. Such reasons include **measurement errors** and **outliers** .

**Measurement error** is the amount you are off by reading a ruler or graph.

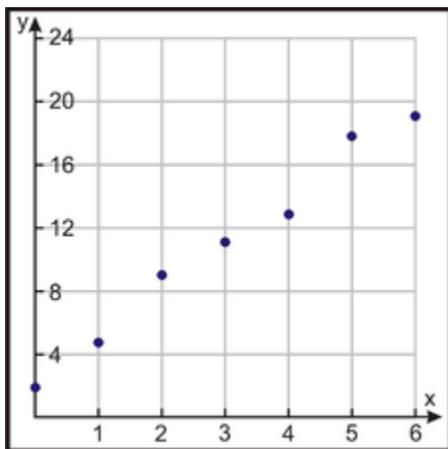
An **outlier** is a data point that does not fit with the general pattern of the data. It tends to be “outside” the majority of the scatter plot.

## Example A

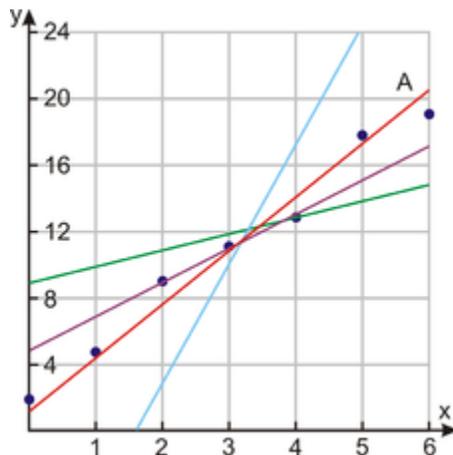
Make a scatter plot of the following ordered pairs.

(0, 2), (1, 4.5), (2, 9), (3, 11), (4, 13), (5, 18), (6, 19.5)

Solution: Graph each ordered pair on one Cartesian plane.



Notice that the points graphed on the plane above look like they might be part of a straight line, although they would not fit perfectly. If the points were perfectly lined up, it would be quite easy to draw a line through all of them and find the equation of that line. However, if the points are “scattered,” we try to find a line that best fits the data. The graph below shows several potential **lines of best fit** .



You see that we can draw many lines through the points in our data set. These lines have equations that are very different from each other. We want to use the line that is

closest to **all** the points on the graph. The best candidate in our graph is the red line,  $A$ . Line  $A$  is the **line of best fit** for this scatter plot.

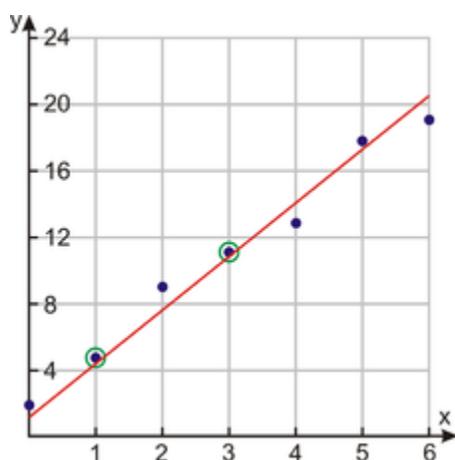
### Writing Equations for Lines of Best Fit

Once you have decided upon your line of best fit, you need to write its equation by finding two points on it and using either:

- Point-[slope](#) form;
- Standard form; or
- Slope-intercept form.

The form you use will depend upon the situation and the ease of finding the  $y$ -intercept.

Using the red line from the example above, locate two points on the line.



Find the [slope](#):  $m = \frac{11-4.5}{3-1} = \frac{6.5}{2} = 3.25$ .

Then  $y = 3.25x + b$ .

Substitute (3, 11) into the equation.  $11 = 3.25(3) + b \Rightarrow b = 1.25$

The equation for the line that fits the data best is  $y = 3.25x + 1.25$ .

### Finding Equations for Lines of Best Fit Using a Calculator

Graphing calculators can make writing equations of best fit easier and more accurate. Two people working with the same data might get two different equations because they would be drawing different lines. To get the most accurate equation for the line, we can use a graphing calculator. The calculator uses a mathematical algorithm to find the line that minimizes the error between the data points and the line of best fit.

### Example B

Use a graphing calculator to find the equation of the line of best fit for the following data: (3, 12), (8, 20), (1, 7), (10, 23), (5, 18), (8, 24), (11, 30), (2, 10).

**Solution:**

**Step 1:** Input the data into your calculator . Press **[STAT]** and choose the **[EDIT]** option.

| L1         | L2    | L3 | 2 |
|------------|-------|----|---|
| 1          | 7     |    |   |
| 10         | 23    |    |   |
| 5          | 18    |    |   |
| 8          | 24    |    |   |
| 11         | 30    |    |   |
| 2          | 10    |    |   |
| -----      | ----- |    |   |
| L2(8) = 10 |       |    |   |

Input the data into the table by entering the  $x$  values in the first column and the  $y$  values in the second column.

**Step 2:** Find the equation of the line of best fit.

| EDIT           | TESTS |
|----------------|-------|
| 1:1-Var Stats  |       |
| 2:2-Var Stats  |       |
| 3:Med-Med      |       |
| 4:LinReg(ax+b) |       |
| 5:QuadReg      |       |
| 6:CubicReg     |       |
| 7:QuartReg     |       |

Press **[STAT]** again and use the right arrow to select **[CALC]** at the top of the screen.

Choose option number 4:  $LinReg(ax + b)$  and press **[ENTER]** . The calculator will display  $LinReg(ax + b)$  .

|              |
|--------------|
| LinReg       |
| $y = ax + b$ |
| $a = 2.01$   |
| $b = 5.94$   |

Press **[ENTER]** and you will be given the  $a$  and  $b$  values.

Here  $a$  represents the slope and  $b$  represents the  $y$ - intercept of the equation. The linear regression line is  $y = 2.01x + 5.94$  .

**Step 3:** Draw the scatter plot.

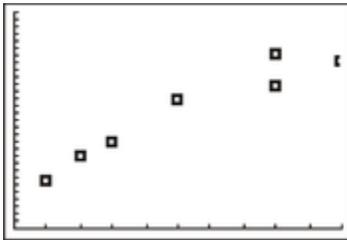
| Plot1     | Plot2 | Plot3 |
|-----------|-------|-------|
| Off       |       |       |
| Type: [ ] | [ ]   | [ ]   |
| Xlist: L1 |       |       |
| Ylist: L2 |       |       |
| Mark: [ ] |       |       |

To draw the scatter plot press **[STATPLOT]** **[2nd]** **[Y=]** .

Choose Plot 1 and press **[ENTER]** .

Press the **On** option and choose the Type as scatter plot (the one highlighted in black).

Make sure that the  $X$  list and  $Y$  list names match the names of the columns of the table in Step 1.



Choose the box or plus as the mark since the simple dot may make it difficult to see the points.

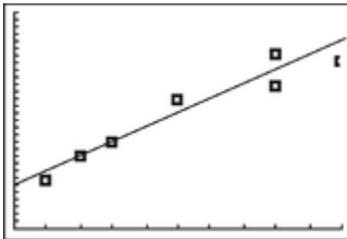
Press **[GRAPH]** and adjust the window size so you can see all the points in the scatter plot.

**Step 4:** Draw the line of best fit through the scatter plot.

Press **[Y=]** .

Enter the equation of the line of best fit that you just found:  $Y_1 = 2.01X + 5.94$  .

Press **[GRAPH]** .



Using Lines of Best Fit to Solve Situations

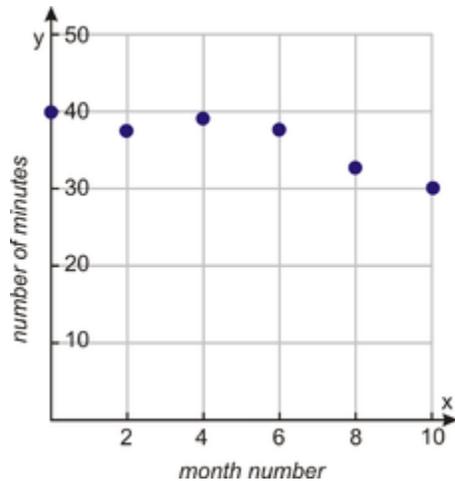
### **Example C**

Gal is training for a 5K race (a total of 5000 meters, or about 3.1 miles). The following table shows her times for each month of her training program. Assume here that her times will decrease in a straight line with time. Find an equation of a line of fit. Predict her running time if her race is in August.

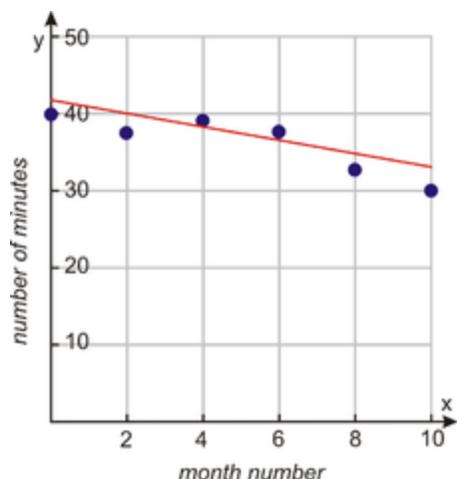


| Month    | Month number | Average time (minutes) |
|----------|--------------|------------------------|
| January  | 0            | 40                     |
| February | 1            | 38                     |
| March    | 2            | 39                     |
| April    | 3            | 38                     |
| May      | 4            | 33                     |
| June     | 5            | 30                     |

Solution: Begin by making a scatter plot of Gal's running times. The independent variable,  $x$ , is the month number and the dependent variable,  $y$ , is the running time in minutes. Plot all the points in the table on the coordinate plane.



Draw a line of fit. When doing this by eye, there are many lines that look like a good fit, so you just have to use your best judgement.



Choose two points on the line you chose: (0, 41) and (4, 38).

Find the equation of the line, first noticing that one of our points, (0,41), is the  $y$ -intercept. Now, all we need is to find the slope.

$$m = \frac{38 - 41}{4 - 0} = -\frac{3}{4}$$

$$y = -\frac{3}{4}x + 41$$

In a real-world problem, the slope and  $y$ -intercept have a physical significance.

$$\text{Slope} = \frac{\text{number of minutes}}{\text{month}}$$

Since the slope is negative, the number of minutes Gal spends running a 5K race decreases as the months pass. The slope tells us that Gal's running time decreases 0.75 minutes per month.

The  $y$ -intercept tells us that when Gal started training, she ran a distance of 5K in 41 minutes, which is just an estimate, since the actual time was 40 minutes.

The problem asks us to predict Gal's running time in August. Since June is assigned to month number five, August will be month number seven. Substitute  $x = 7$  into the line-of-best-fit-equation.

$$y = -\frac{3}{4}(7) + 41 = -\frac{21}{4} + 41 = -\frac{21}{4} + \frac{164}{4} = \frac{143}{4} = 35\frac{3}{4}$$

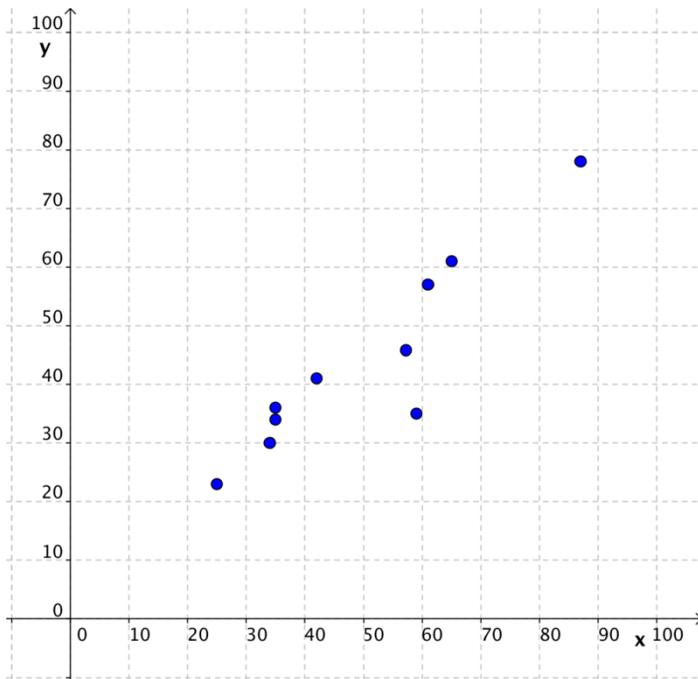
The equation predicts that Gal will be running the 5K race in 35.75 minutes.

## Guided Practice

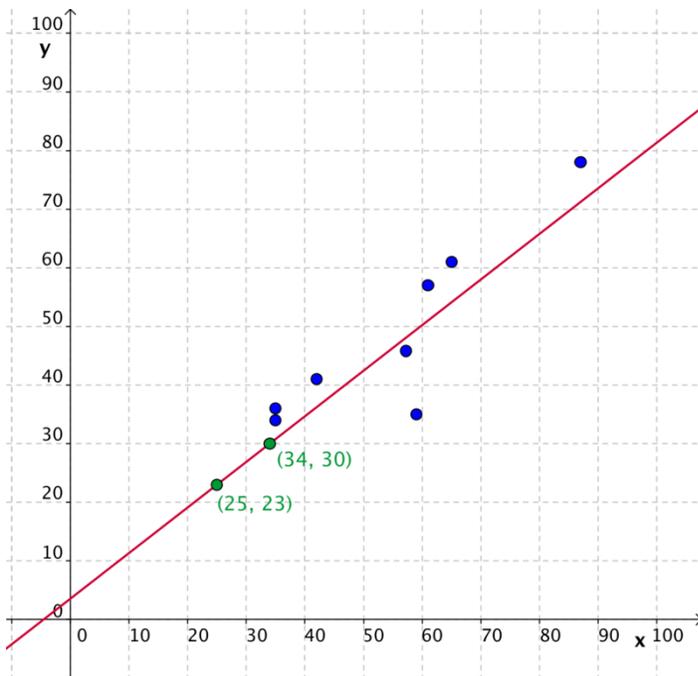
Make a scatter plot and find the equation of a best fit line for the following set of points: (57, 45) (65, 61) (34, 30) (87, 78) (42, 41) (35, 36) (59, 35) (61, 57) (25, 23) (35, 34).

## Solution:

First we will make a scatter plot:



Next, draw in a line, finding the best fit by eye:



Since the two green points, (34,30) and (25,23), are on the line, we can use them to write the equation. First, find the slope:

$$m = \frac{30-23}{34-25} = \frac{7}{9}$$

Plugging this into point-slope:

$$y - 30 = \frac{7}{9}(x - 34)$$

$$y - 30 = \frac{7}{9}x - \frac{7}{9} \cdot (34)$$

$$\text{Since } \frac{7}{9} \cdot (34) \approx 26.44$$

$$y - 30 = \frac{7}{9}x - 26.44$$

$$y - 30 + 30 = \frac{7}{9}x - 26.44 + 30$$

$$y = \frac{7}{9}x + 3.56$$

---

## Practice

Sample explanations for some of the practice exercises below are available by viewing the following video. Note that there is not always a match between the number of the practice exercise in the video and the number of the practice exercise listed in the following exercise set. However, the practice exercise is the same in both. [CK-12 Basic Algebra: Fitting a Line to Data](#)(7:48)

1. What is a *scatter plot*? How is this different from other graphs you have created?
2. Define *line of best fit*.
3. What is an *outlier*? How can an outlier be spotted on a graph?
4. What are the two methods of finding a line of best fit?
5. Explain the steps needed to find a line of best fit “by hand.” What are some problems with using this method?

For each data set, draw the scatter plot and find the equation of the line of best fit by hand.

6. (32, 43) (54, 61) (89, 94) (25, 34) (43, 56) (58, 67) (38, 46) (47, 56) (39, 48)

7. (12, 18) (5, 24) (15, 16) (11, 19) (9, 12) (7, 13) (6, 17) (12, 14)

8. (3, 12) (8, 20) (1, 7) (10, 23) (5, 18) (8, 24) (2, 10)

In 9 – 11, for each data set, use a graphing calculator to find the equation of the line of best fit.

9. (57, 45) (65, 61) (34, 30) (87, 78) (42, 41) (35, 36) (59, 35) (61, 57) (25, 23) (35, 34)

10. (32, 43) (54, 61) (89, 94) (25, 34) (43, 56) (58, 67) (38, 46) (47, 56) (95, 105) (39, 48)

11. (12, 18) (3, 26) (5, 24) (15, 16) (11, 19) (0, 27) (9, 12) (7, 13) (6, 17) (12, 14)

12. Shiva is trying to beat the samosa eating record. The current record is 53.5 samosas in 12 minutes. The following table shows how many samosas he eats during his daily practice for the first week of his training. Will he be ready for the contest if it occurs two weeks from the day he started training? What are the meanings of the slope and the  $y$ -intercept in this problem?

| Day | No. of Samosas |
|-----|----------------|
| 1   | 30             |
| 2   | 34             |
| 3   | 36             |
| 4   | 36             |
| 5   | 40             |
| 6   | 43             |
| 7   | 45             |

13. Nitisha is trying to find the elasticity coefficient of a Superball. She drops the ball from different heights and measures the maximum height of the resulting bounce. The table below shows her data. Draw a scatter plot and find the equation. What is the initial height if the bounce height is 65 cm? What are the meanings of the slope and the  $y$ -intercept in this problem?

| Initial height (cm) | Bounce height (cm) |
|---------------------|--------------------|
| 30                  | 22                 |
| 35                  | 26                 |
| 40                  | 29                 |
| 45                  | 34                 |
| 50                  | 38                 |
| 55                  | 40                 |

| Initial height (cm) | Bounce height (cm) |
|---------------------|--------------------|
| 60                  | 45                 |
| 65                  | 50                 |
| 70                  | 52                 |

14. Baris is testing the burning time of “BriteGlo” candles. The following table shows how long it takes to burn candles of different weights. Let's assume it's a linear relation. We can then use a line to fit the data. If a candle burns for 95 hours, what must be its weight in ounces?

| Candle Burning Time Based on Candle Weight |              |
|--|--------------|
| Candle weight (oz)                         | Time (hours) |
| 2  | 15           |
| 3  | 20           |
| 4  | 35           |
| 5  | 36           |
| 10   | 80           |
| 16   | 100          |
| 22   | 120          |
| 26   | 180          |

15. The table below shows the [median](#) California family income from 1995 to 2002 as reported by the U.S. Census Bureau. Draw a scatter plot and find the equation. What would you expect the median annual income of a Californian family to be in year 2010? What are the meanings of the slope and the  $y$ -intercept in this problem?

| Year | Income |
|------|--------|
| 1995 | 53,807 |
| 1996 | 55,217 |
| 1997 | 55,209 |
| 1998 | 55,415 |
| 1999 | 63,100 |
| 2000 | 63,206 |
| 2001 | 63,761 |
| 2002 | 65,766 |

### Mixed Review

16. Sheri bought an espresso machine and paid \$119.64 including tax. The sticker price was \$110.27. What was the percent of tax?
17. What are the means of  $\frac{4}{x} = \frac{141}{98}$ ? What are the extremes?
18. Solve the proportion in question 17.
19. The distance traveled varies directly with the time traveled. If a car has traveled 328.5 miles in 7.3 hours, how many hours will it take to travel 82.8 miles?
20. Evaluate  $t(x) = 0.85x$  when  $x = 6015$ .