Lesson 10: Linear Models

Student Outcomes

- Students identify situations where it is reasonable to use a linear function to model the relationship between two numerical variables.
- Students interpret slope and the initial value in a data context.

Lesson Notes

In previous lessons, students were given a set of bivariate data on variables that were linearly related. Students constructed a scatter plot of the data, informally fit a line to the data, and found the equation of their prediction line. The lessons also discussed criteria students could use to determine what might be considered the *best* fitting prediction line for a given set of data. A more formal discussion of this topic occurs in Algebra I.

This lesson introduces a formal statistical terminology for the two variables that define a bivariate data set. In a prediction context, we refer to the x-variable as the independent variable, explanatory variable, or predictor variable. We refer to the y-variable as the dependent variable, response variable, or predicted variable. Students should become equally comfortable with using the pairings (independent, dependent), (explanatory, response), and (predictor, predicted). Statistics builds on data, and in this lesson, students investigate bivariate data that are linearly related. Students examine how the dependent variable relates to the independent variable or how the predicted variable relates to the predictor variable. Students also need to connect the linear function in words to a symbolic form that represents a linear function. In most cases, the independent variable is denoted by x and the dependent variable by y.

Similar to lessons at the beginning of this module, this lesson works with *exact* linear relationships. This is done to build conceptual understanding of how structural elements of the modeling equation are explained in context. Students will apply this thinking to more authentic data contexts in the next lesson.

Classwork

In previous lessons, you used data that follow a linear trend either in the positive direction or the negative direction and informally fitted a line through the data. You determined the equation of an informal fitted line and used it to make predictions.

In this lesson, you will use a function to model a linear relationship between two numerical variables and interpret the slope and intercept of the linear model in the context of the data. Recall that a function is a rule that relates a dependent variable to an independent variable.

In statistics, a dependent variable is also called a response variable or a predicted variable. An independent variable is also called an explanatory variable or a predictor variable.

Scaffolding:

- A dependent variable is also called a response or predicted variable.
- An independent variable is also called an explanatory or predictor variable.
- It is important to make the interchangeability of these terms clear to ELL students.
- For each of the pairings, students should have the chance to read, write, speak, and hear them on multiple occasions.

Example 1 (5 minutes)

This lesson begins by challenging students' understanding of the terminology. Read through the opening text and explain the difference between dependent and independent variables. Pose the question to the class at the end of the example, and allow for multiple responses.

- What are some other possible numerical independent variables that could relate to how well you are going to do on the quiz?
  - *How many hours of sleep I got the night before.*

Example 1

Predicting the value of a numerical dependent (response) variable based on the value of a given numerical independent variable has many applications in statistics. The first step in the process is to identify the dependent (predicted) variable and the independent (predictor) variable.

There may be several independent variables that might be used to predict a given dependent variable. For example, suppose you want to predict how well you are going to do on an upcoming statistics quiz. One possible independent variable is how much time you spent studying for the quiz. What are some other possible numerical independent variables that could relate to how well you are going to do on the quiz?

Exercise 1 (5 minutes)

Exercise 1 requires students to write two possible explanatory variables that might be used for each of several given response variables. Give students a moment to think about each response variable, and then discuss the answers as a class. Allow for multiple student responses.

Exercises 1–2

1. **For each of the following dependent (response) variables, identify two possible numerical independent (explanatory) variables that might be used to predict the value of the dependent variable.**

   Answers will vary. Here again, make sure that students are defining their explanatory variables (predictors) correctly and that they are numerical.

Response Variable	Possible Explanatory Variables
Height of a son	*1. Height of the boy's father* *2. Height of the boy's mother*
Number of points scored in a game by a basketball player	*1. Number of shots taken in the game* *2. Number of minutes played in the game*
Number of hamburgers to make for a family picnic	*1. Number of people in the family* *2. Price of hamburger meat*
Time it takes a person to run a mile	*1. Height above sea level of track field* *2. Number of practice days*
Amount of money won by a contestant on Jeopardy! (television game show)	*1. IQ of the contestant* *2. Number of questions correctly answered*
Fuel efficiency (in miles per gallon) for a car	*1. Weight of the car* *2. Size of the car's engine*
Number of honey bees in a beehive at a particular time	*1. Size of a queen bee* *2. Amount of honey harvested from the hive*
Number of blooms on a dahlia plant	*1. Amount of fertilizer applied to the plant* *2. Amount of water applied to the plant*
Number of forest fires in a state during a particular year	*1. Number of acres of forest in the state* *2. Amount of rain in the state that year*

Exercise 2 (5 minutes)

This exercise reverses the format and asks students to provide a response variable for each of several given explanatory variables. Again, give students a moment to consider each independent variable. Then, discuss the dependent variables as a class. Allow for multiple student responses.

2. **Now, reverse your thinking. For each of the following numerical independent variables, write a possible numerical dependent variable.**

Dependent Variable	Possible Independent Variables
Time it takes a student to run a mile	Age of a student
Distance a golfer drives a ball from a tee	Height of a golfer
Time it takes pain to disappear	Amount of a pain-reliever taken
Amount of money a person makes in a lifetime	Number of years of education
Number of tomatoes harvested in a season	Amount of fertilizer used on a garden
Price of a diamond ring	Size of a diamond in a ring
A baseball team's batting average	Total salary for all of a team's players

Example 2 (3–5 minutes)

This example begins the study of an exact linear relationship between two numerical variables. Example 2 and Exercises 3–9 address bivariate data that have an exact functional form, namely linear. Students become familiar with an equation of the form: $y = \text{intercept} + (\text{slope})x$. They connect this representation to the equation of a linear function ($y = mx + b$ or $y = a + bx$) developed in previous modules. Make sure students clearly identify the slope and the y-intercept as they describe a linear function. Students interpret slope as the change in the dependent variable (the y-variable) for an increase of one unit in the independent variable (the x-variable).

For example, if $\text{exam score} = 57 + 8(\text{study time})$, or equivalently $y = 57 + 8x$, where y represents the exam score and x represents the study time in hours, then an increase of one hour in study time produces an increase of 8 points in the predicted exam score. Encourage students to interpret slope in the context of the problem. Their interpretation of slope as simply "rise over run" is not sufficient in a statistical setting.

Students should become comfortable writing linear models using descriptive words (such as exam score and study time) or using symbols, such as x and y, to represent variables. Using descriptive words when writing model equations can help students keep the context in mind, which is important in statistics.

Note that bivariate numerical data that do not have an exact linear functional form but do have a linear trend are covered in the next lesson. Starting with Example 2, this lesson covers only contexts in which the linear relationship is exact.

Give students a moment to read through Example 2. For ELL students, consider reading the example aloud.

Example 2

A cell-phone company offers the following basic cell-phone plan to its customers: A customer pays a monthly fee of $\$40.00$. In addition, the customer pays $\$0.15$ per text message sent from the cell phone. There is no limit to the number of text messages per month that could be sent, and there is no charge for receiving text messages.

Exercises 3–9 (10–15 minutes)

These exercises build on earlier lessons in Module 6. Provide time for students to develop answers to the exercises. Then, confirm their answers as a class.

Exercises 3–11

3. **Determine the following:**
   a. **Justin never sends a text message. What would be his total monthly cost?**

      Justin's monthly cost would be $\$40.00$***.***

   b. **During a typical month, Abbey sends 25 text messages. What is her total cost for a typical month?**

      Abbey's monthly cost would be $\$40.00 + \$0.15(25)$***, or*** $\$43.75$***.***

   c. **Robert sends at least 250 text messages a month. What would be an estimate of the least his total monthly cost is likely to be?**

      Robert's monthly cost would be $\$40.00 + \$0.15(250)$***, or*** $\$77.50$***.***

Scaffolding:

Using a table may help students better understand the relationship between the number of text messages and total monthly cost.

Number of messages	Total Cost ($)
0	$40 + 0.15(0) = 40$
1	$40 + 0.15(1) = 40.15$
2	$40 + 0.15(2) = 40.30$
3	$40 + 0.15(3) = 40.45$
4	$40 + 0.15(4) = 40.60$
5	$40 + 0.15(5) = 40.75$
10	$40 + 0.15(10) = 41.50$

There is a cost increase of $0.15 for every additional text message sent from the phone.

4. **Use descriptive words to write a linear model describing the relationship between the number of text messages sent and the total monthly cost.**

   $\textbf{Total monthly cost} = \$40.00 + (\textbf{number of text messages}) \cdot \$0.15$

5. **Is the relationship between the number of text messages sent and the total monthly cost linear? Explain your answer.**

   Yes, for each text message, the total monthly cost goes up by $\$0.15$***. From our previous work with linear functions, this would indicate a linear relationship.***

MP.4

6. **Let x represent the independent variable and y represent the dependent variable. Use the variables x and y to write the function representing the relationship you indicated in Exercise 4.**

   Students show the process in developing a model of the relationship between the two variables.
   $y = 0.15x + 40$ ***or*** $y = 40 + 0.15x$

7. **Explain what $\$0.15$ represents in this relationship.**

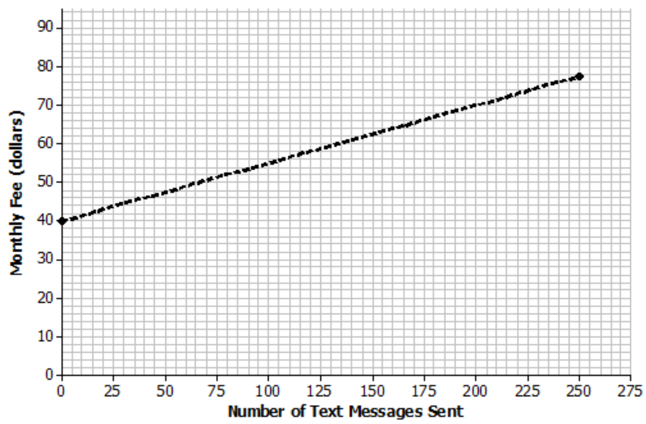
   $\$0.15$ ***represents the slope of the linear relationship or the change in the total monthly cost is*** $\$0.15$ ***for an increase of one text message. (Students need to clearly explain that slope is the change in the dependent variable for a*** 1***-unit increase in the independent variable.)***

8. **Explain what $\$40.00$ represents in this relationship.**

   $\$40.00$ ***represents the fixed monthly fee or the*** y***-intercept of this relationship. This is the value of the total monthly cost when the number of text messages is*** 0***.***

9. Sketch a graph of this relationship on the following coordinate grid. Clearly label the axes and include units in the labels.

Anticipated response: Students label the x-axis as the number of text messages. They label the y-axis as the total monthly cost. Students use any two points they derived in Exercise 3. The following graph uses the point of $(0, 40)$ for Justin and the point $(250, 77.5)$ for Robert. Highlight the intercept of $(0, 40)$, along with the slope of the line they sketched. Also, point out that the line the student draws should be a dotted line (and not a solid line). The number of text messages can only be whole numbers, and as a result, the line representing this relationship should indicate that values in between the whole numbers representing the text messages are not part of the data.



Exercise 10 (5 minutes)

If time is running short, teachers may want to choose either Exercise 10 or 11 to develop in class and assign the other to the Problem Set. Let students continue to work with a partner, and confirm answers as a class.

10. LaMoyne needs four more pieces of lumber for his scout project. The pieces can be cut from one large piece of lumber according to the following pattern.

The lumberyard will make the cuts for LaMoyne at a fixed cost of $\$2.25$ plus an additional cost of 25 cents per cut. One cut is free.

a. What is the functional relationship between the total cost of cutting a piece of lumber and the number of cuts required? What is the equation of this function? Be sure to define the variables in the context of this problem.

As students uncover the information in this problem, they should realize that the functional relationship between the total cost and number of cuts is linear. Noting that one cut is free, the equation could be written in one of the following ways:

$\text{Total cost for cutting} = 2.25 + (0.25)(\text{number of cuts} - 1)$
$y = 2.25 + (0.25)(x - 1)$, *where* $x = \text{number of cuts.}$

$\text{Total cost for cutting} = 2 + (0.25)(\text{number of cuts})$
$y = 2 + 0.25x$, *where* $x = \text{number of cuts.}$

$\text{Total cost for cutting} = 2.25 + (0.25)(\text{number of paid cuts})$
$y = 2.25 + 0.25x$, *where* $x = \text{number of paid cuts.}$

b. **Use the equation to determine LaMoyne's total cost for cutting.**

LaMoyne requires three cuts, one of which is free. Using any of the three forms given in part (a) yields a total cost for cutting of $\$2.75$.

c. **In the context of this problem, interpret the slope of the equation in words.**

Using any of the three forms, each additional cut beyond the free one adds $\$0.25$ *to the total cost for cutting.*

d. **Interpret the y-intercept of your equation in words in the context of this problem. Does interpreting the intercept make sense in this problem? Explain.**

If no cuts are required, then there is no fixed cost for cutting. So, it does not make sense to interpret the intercept in the context of this problem.

Exercise 11 (5–7 minutes)

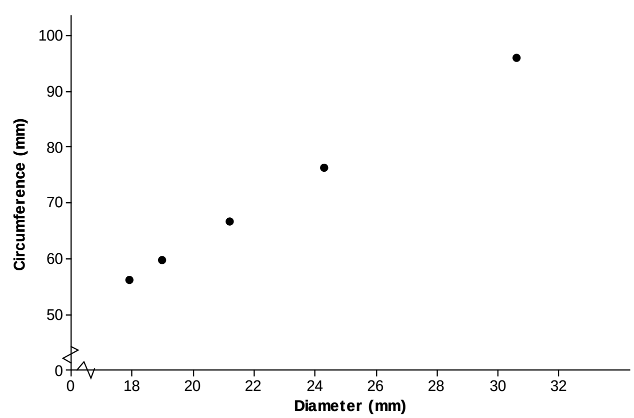
Let students work with a partner. Then, confirm answers as a class.

11. **Omar and Olivia were curious about the size of coins. They measured the diameter and circumference of several coins and found the following data.**

US Coin	Diameter (mm)	Circumference (mm)
Penny	19.0	59.7
Nickel	21.2	66.6
Dime	17.9	56.2
Quarter	24.3	76.3
Half Dollar	30.6	96.1

a. **Wondering if there was any relationship between diameter and circumference, they thought about drawing a picture. Draw a scatter plot that displays circumference in terms of diameter.**

Students may need some help in deciding which is the independent variable and which is the dependent variable. Hopefully, they have seen from previous problems that whenever one variable, say variable A, is to be expressed in terms of some variable B, then variable A is the dependent variable and variable B is the independent variable. So, circumference is being taken as the dependent variable in this problem, and diameter is being taken as the independent variable.



b. **Do you think that circumference and diameter are related? Explain.**

You may need to point out to students that because the data are rounded to one decimal place, the points on the scatter plot may not fall exactly on a line; however, they should. Circumference and diameter are linearly related.

c. **Find the equation of the function relating circumference to the diameter of a coin.**

Again, because of rounding error, equations that students find may be slightly different depending on which points they choose to do their calculations. Hopefully, they all arrive at something close to a circumference equal to 3.14 (pi) multiplied by diameter.

For example, the slope between $(19, 59.7)$ and $(30.6, 96.1)$ is $\frac{96.1-59.7}{30.6-19} = 3.1379$, which rounds to 3.14.

The intercept may be found using $59.7 = a + (3.14)(19.0)$, which yields $a = 0.04$, which rounds to 0.

Therefore, $C = 3.14d + 0 = 3.14d$.

d. **The value of the slope is approximately equal to the value of π. Explain why this makes sense.**

The slope is identified as pi. (Note: Most students have previously studied the relationship between circumference and diameter of a circle. However, if students have not yet seen this result, you can discuss the interesting result that if the circumference of a circle is divided by its diameter, the result is a constant, namely 3.14 rounded to two decimal places, no matter what circle is being considered.)

e. **What is the value of the y-intercept? Explain why this makes sense.**

If the diameter of a circle is 0 (a point), then according to the equation, its circumference is 0. That is true, so interpreting the intercept of 0 makes sense in this problem.

Closing (2–3 minutes)

- Think back to Exercise 10. If the equation that models LaMoyne's total cost of cutting is given by $y = 2.25 + 0.25x$, what are the dependent and independent variables?
  - *Independent variable is the number of paid cuts. Dependent variable is the total cost for cutting.*
- What are the meanings of the y-intercept and slope in context?
  - *The y-intercept is the fee for the first cut; however, if no cuts are required then there is no fixed cost for cutting. The slope is the cost per cut after the first.*
- How are these examples different from the data we have been studying before this lesson?
  - *These examples are exact linear relationships.*

Lesson Summary

- **A linear functional relationship between a dependent and independent numerical variable has the form $y = mx + b$ or $y = a + bx$.**
- **In statistics, a dependent variable is one that is predicted and an independent variable is the one that is used to make the prediction.**
- **The graph of a linear function describing the relationship between two variables is a line.**

Exit Ticket (5–7 minutes)

Name ____ Date____________________

Lesson 10: Linear Models

Exit Ticket

Suppose that a cell-phone monthly rate plan costs the user 5 cents per minute beyond a fixed monthly fee of $\$20$. This implies that the relationship between monthly cost and monthly number of minutes is linear.

1. Write an equation in words that relates total monthly cost to monthly minutes used. Explain how you found your answer.

2. Write an equation in symbols that relates the total monthly cost (y) to monthly minutes used (x).

3. What is the cost for a month in which 182 minutes are used? Express your answer in words in the context of this problem.

Exit Ticket Sample Solutions

Suppose that a cell-phone monthly rate plan costs the user 5 cents per minute beyond a fixed monthly fee of $\$20$. This implies that the relationship between monthly cost and monthly number of minutes is linear.

1. **Write an equation in words that relates total monthly cost to monthly minutes used. Explain how you arrived at your answer.**

   The equation is given by total monthly cost $= 20 + 0.05$ (number of minutes used for a month).

   The y-intercept in the equation is the fixed monthly cost, $\$20$.

   The slope is the amount paid per minute of cell phone usage, or $\$0.05$ per minute.

   The linear form is total monthly cost $=$ fixed cost $+$ cost per minute (number of minutes used for a month).

2. **Write an equation in symbols that relates the total monthly cost (y) to monthly minutes used (x).**

   The equation is $y = 20 + 0.05x$, where y is the total cost for a month in dollars and x is cell phone usage for the month in minutes.

3. **What is the cost for a month in which 182 minutes are used? Express your answer in words in the context of this problem.**

   The total monthly cost in a month using 182 minutes would be 20 dollars $+$ $(0.05$ dollars per minute$)(182$ minutes$) = \$29.10$.

   Be sure students pay attention to the meanings of the units, noting that units on one side of the equation must be the same as units on the other side.

Problem Set Sample Solutions

1. **The Mathematics Club at your school is having a meeting. The advisor decides to bring bagels and his award-winning strawberry cream cheese. To determine his cost, from past experience he figures 1.5 bagels per student. A bagel costs 65 cents, and the special cream cheese costs $\$3.85$ and will be able to serve all of the anticipated students attending the meeting.**

   a. **Find an equation that relates his total cost to the number of students he thinks will attend the meeting.**

      Encourage students to write a problem in words in its context. For example, the advisor's total cost $=$ cream cheese fixed cost $+$ cost of bagels. The cost of bagels depends on the unit cost of a bagel times the number of bagels per student times the number of students. So, with symbols, if c denotes the total cost in dollars and n denotes the number of students, then $c = 3.85 + (0.65)(1.5)(n)$, or $c = 3.85 + 0.975n$.

   b. **In the context of the problem, interpret the slope of the equation in words.**

      For each additional student, the cost goes up by 0.975 dollars, or 97.5 cents.

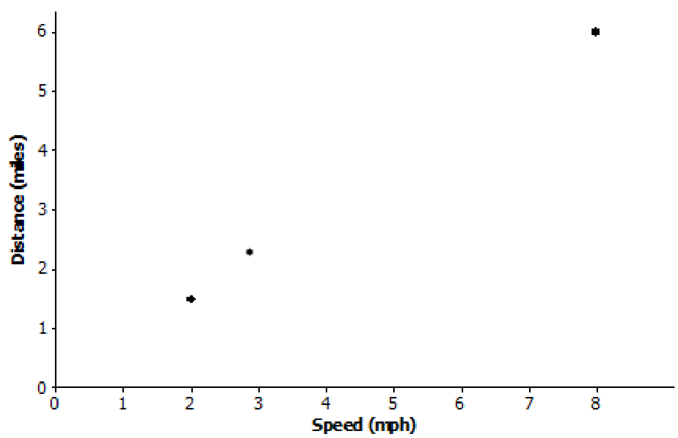
   c. **In the context of the problem, interpret the y-intercept of the equation in words. Does interpreting the intercept make sense? Explain.**

      If there are no students, the total cost is $\$3.85$. Students could interpret this by saying that the meeting was called off before any bagels were bought, but the advisor had already made his award-winning cream cheese, so the cost is $\$3.85$. The intercept makes sense. Other students might argue otherwise.

2. John, Dawn, and Ron, get together to exercise (walk/jog) for 45 minutes. John has arthritic knees but manages to walk $1\frac{1}{2}$ miles. Dawn walks $2\frac{1}{4}$ miles, while Ron manages to jog 6 miles.

   a. Draw an appropriate graph and connect the points to show that there is a linear relationship between the distance that each traveled based on how fast each traveled (speed). Note that the speed for a person who travels 3 miles in 45 minutes, or $\frac{3}{4}$ hours, is $3 \div \frac{3}{4} = 4$ miles per hour.

   John's speed is $\left(1\frac{1}{2} \div \frac{3}{4}\right) = 2$ miles per hour, Dawn's speed is $2\frac{1}{4} \div \frac{3}{4} = 3$ miles per hour, and Ron's speed is $6 \div \frac{3}{4} = 8$ miles per hour. Students may draw the scatter plot incorrectly. Note that distance is to be expressed in terms of speed so that distance is the dependent variable on the vertical axis, and speed is the independent variable on the horizontal axis.



   b. Find an equation that expresses distance in terms of speed (how fast one goes).

   The slope is $\frac{6-1.5}{8-2} = 0.75$, so the equation of the line through these points is distance $= a + (0.75)(\text{speed})$.

   Next, find the intercept. For example, solve $6 = a + (0.75)(8)$ for a, which yields $a = 0$.

   So, the equation is distance $= 0.75(\text{speed})$.

   c. In the context of the problem, interpret the slope of the equation in words.

   If someone increases his or her speed by 1 mile per hour, then that person travels 0.75 additional miles in 45 minutes.

   d. In the context of the problem, interpret the y-intercept of the equation in words. Does interpreting the intercept make sense? Explain.

   The intercept of 0 makes sense because if the speed is 0 miles per hour, then the person is not moving. So, the person travels no distance.

Note: Simple interest is developed in the next two problems. It is an excellent example of an application of a linear function. If students have not worked previously with finance problems of this type, then you may need to carefully explain simple interest as stated in the problem. It is an important discussion to have with students if time permits. If this discussion is not possible and students have not worked previously with any financial applications, then omit these problems.

3. Simple interest is money that is paid on a loan. Simple interest is calculated by taking the amount of the loan and multiplying it by the rate of interest per year and the number of years the loan is outstanding. For college, Jodie's older brother has taken out a student loan for $\$4,500$ at an interest rate of 5.6%, or 0.056. When he graduates in four years, he will have to pay back the loan amount plus interest for four years. Jodie is curious as to how much her brother will have to pay.

   a. Jodie claims that his brother will have to pay a total of $\$5,508$. Do you agree? Explain. As an example, 8% simple interest on $\$1,200$ for one year is $(0.08)(1200) = \$96$. The interest for two years would be $2 \times \$96$, or $\$192$.

   The total cost to repay $=$ amount of loan $+$ interest on the loan.

   Interest on the loan is the amount of simple interest for one year times the number of years the loan is outstanding.

   The annual simple interest amount is $(0.056)(\$4500) = \$252$ per year.

   For four years, the interest amount is $4(\$252) = \$1008$.

   So, the total cost to repay the loan is $\$4500 + \$1008 = \$5508$. Jodie is right.

   b. Write an equation for the total cost to repay a loan of $\$P$ if the rate of interest for a year is r (expressed as a decimal) for a time span of t years.

   Note: Work with students in identifying variables to represent the values discussed in this exercise. For example, the total cost to repay a loan is $P +$ the amount of interest on P for t years, or $P + I$, where $I =$ interest.

   The amount of interest per year is P times the annual interest. Let r represent the interest rate per year as a decimal.

   The total amount of simple interest for t years is rt, where r is the annual rate as a decimal (e.g., 5% is 0.05).

   So, if c denotes the total cost to repay the loan, then $c = P + (rt)P$.

   c. If P and r are known, is the equation a linear equation?

   If P and r are known, then the equation should be written as $c = P + (rP)t$, which is the linear form where c is the dependent variable and t is the independent variable.

   d. In the context of this problem, interpret the slope of the equation in words.

   For each additional year that the loan is outstanding, the total cost to repay the loan is increased by $\$rP$.

   As an example, consider Jodie's brother's equation for t years: $c = 4500 + (0.056)(4500)t$, or $c = 4500 + 252t$. For each additional year that the loan is not paid off, the total cost increases by $\$252$.

   e. In the context of this problem, interpret the intercept of the equation in words. Does interpreting the intercept make sense? Explain.

   The 0 value of time t means at the time the loan was taken out. At that time, no interest has been accumulated, so the intercept of $\$4,500$ as the cost to repay the loan after 0 years makes sense.