Chapter 1: Functions

Section 1.1 Functions and Function Notation

What is a Function?
The natural world is full of relationships between quantities that change. When we see these relationships, it is natural for us to ask "If I know one quantity, can I then determine the other?" This establishes the idea of an input quantity, or independent variable, and a corresponding output quantity, or dependent variable. From this we get the notion of a functional relationship in which the output can be determined from the input.

For some quantities, like height and age, there are certainly relationships between these quantities. Given a specific person and any age, it is easy enough to determine their height, but if we tried to reverse that relationship and determine age from a given height, that would be problematic, since most people maintain the same height for many years.

Function
Function: A rule for a relationship between an input, or independent, quantity and an output, or dependent, quantity in which each input value uniquely determines one output value. We say "the output is a function of the input."

Example 1

In the height and age example above, is height a function of age? Is age a function of height?

In the height and age example above, it would be correct to say that height is a function of age, since each age uniquely determines a height. For example, on my 18th birthday, I had exactly one height of 69 inches.

However, age is not a function of height, since one height input might correspond with more than one output age. For example, for an input height of 70 inches, there is more than one output of age since I was 70 inches at the age of 20 and 21.

Example 2

At a coffee shop, the menu consists of items and their prices. Is price a function of the item? Is the item a function of the price?

We could say that price is a function of the item, since each input of an item has one output of a price corresponding to it. We could not say that item is a function of price, since two items might have the same price.

Example 3

In many classes the overall percentage you earn in the course corresponds to a decimal grade point. Is decimal grade a function of percentage? Is percentage a function of decimal grade?

For any percentage earned, there would be a decimal grade associated, so we could say that the decimal grade is a function of percentage. That is, if you input the percentage, your output would be a decimal grade. Percentage may or may not be a function of decimal grade, depending upon the teacher's grading scheme. With some grading systems, there are a range of percentages that correspond to the same decimal grade.

One-to-One Function

Sometimes in a relationship each input corresponds to exactly one output, and every output corresponds to exactly one input. We call this kind of relationship a **one-to-one function**.

From Example 3, *if* each unique percentage corresponds to one unique decimal grade point and each unique decimal grade point corresponds to one unique percentage then it is a one-to-one function.

Try it Now

Let's consider bank account information.

1. Is your balance a function of your bank account number?
   (if you input a bank account number does it make sense that the output is your balance?)

2. Is your bank account number a function of your balance?
   (if you input a balance does it make sense that the output is your bank account number?)

Function Notation

To simplify writing out expressions and equations involving functions, a simplified notation is often used. We also use descriptive variables to help us remember the meaning of the quantities in the problem.

Rather than write "height is a function of age", we could use the descriptive variable h to represent height and we could use the descriptive variable a to represent age.

"height is a function of age"	if we name the function f we write
"h is f of a"	or more simply
$h = f(a)$	we could instead name the function h and write
$h(a)$	which is read "h of a"

Remember we can use any variable to name the function; the notation $h(a)$ shows us that h depends on a. The value "a" must be put into the function "h" to get a result. Be careful - the parentheses indicate that age is input into the function (Note: do not confuse these parentheses with multiplication!).

Function Notation

The notation output = f(input) defines a function named f. This would be read "output is f of input"

Example 4

Introduce function notation to represent a function that takes as input the name of a month, and gives as output the number of days in that month.

The number of days in a month is a function of the name of the month, so if we name the function f, we could write "days = f(month)" or $d = f(m)$. If we simply name the function d, we could write $d(m)$

For example, d(March) = 31, since March has 31 days. The notation $d(m)$ reminds us that the number of days, d (the output) is dependent on the name of the month, m (the input)

Example 5

A function $N = f(y)$ gives the number of police officers, N, in a town in year y. What does $f(2005) = 300$ tell us?

When we read $f(2005) = 300$, we see the input quantity is 2005, which is a value for the input quantity of the function, the year (y). The output value is 300, the number of police officers (N), a value for the output quantity. Remember $N=f(y)$. This tells us that in the year 2005 there were 300 police officers in the town.

Tables as Functions

Functions can be represented in many ways: Words (as we did in the last few examples), tables of values, graphs, or formulas. Represented as a table, we are presented with a list of input and output values.
In some cases, these values represent everything we know about the relationship, while in other cases the table is simply providing us a few select values from a more complete relationship.

Table 1: This table represents the input, number of the month (January = 1, February = 2, and so on) while the output is the number of days in that month. This represents everything we know about the months & days for a given year (that is not a leap year)

(input) Month number, m	1	2	3	4	5	6	7	8	9	10	11	12
(output) Days in month, D	31	28	31	30	31	30	31	31	30	31	30	31

Table 2: The table below defines a function $Q = g(n)$. Remember this notation tells us g is the name of the function that takes the input n and gives the output Q.

n	1	2	3	4	5
Q	8	6	7	6	8

Table 3: This table represents the age of children in years and their corresponding heights. This represents just some of the data available for height and ages of children.

(input) a, age in years	5	5	6	7	8	9	10
(output) h, height inches	40	42	44	47	50	52	54

Example 6

Which of these tables define a function (if any)? Are any of them one-to-one?

Input	Output
2	1
5	3
8	6

Input	Output
-3	5
0	1
4	5

Input	Output
1	0
5	2
5	4

The first and second tables define functions. In both, each input corresponds to exactly one output. The third table does not define a function since the input value of 5 corresponds with two different output values.

Only the first table is one-to-one; it is both a function, and each output corresponds to exactly one input. Although table 2 is a function, because each input corresponds to exactly one output, each output does not correspond to exactly one input so this function is not one-to-one. Table 3 is not even a function and so we don't even need to consider if it is a one-to-one function.

Try it Now

3. If each percentage earned translated to one letter grade, would this be a function? Is it one-to-one?

Solving and Evaluating Functions:

When we work with functions, there are two typical things we do: evaluate and solve. Evaluating a function is what we do when we know an input, and use the function to determine the corresponding output. Evaluating will always produce one result, since each input of a function corresponds to exactly one output.

Solving equations involving a function is what we do when we know an output, and use the function to determine the inputs that would produce that output. Solving a function could produce more than one solution, since different inputs can produce the same output.

Example 7

Using the table shown, where $Q=g(n)$

n	1	2	3	4	5
Q	8	6	7	6	8

a) Evaluate $g(3)$

Evaluating $g(3)$ (read: "g of 3")
means that we need to determine the output value, Q, of the function g given the input value of n=3. Looking at the table, we see the output corresponding to n=3 is Q=7, allowing us to conclude $g(3) = 7$.

b) Solve $g(n) = 6$

Solving $g(n) = 6$ means we need to determine what input values, n, produce an output value of 6. Looking at the table we see there are two solutions: $n = 2$ and $n = 4$.

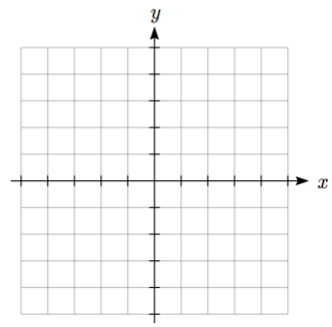
When we input 2 into the function g, our output is $Q = 6$

When we input 4 into the function g, our output is also $Q = 6$

Try it Now

4. Using the function in Example 7, evaluate $g(4)$

Graphs as Functions

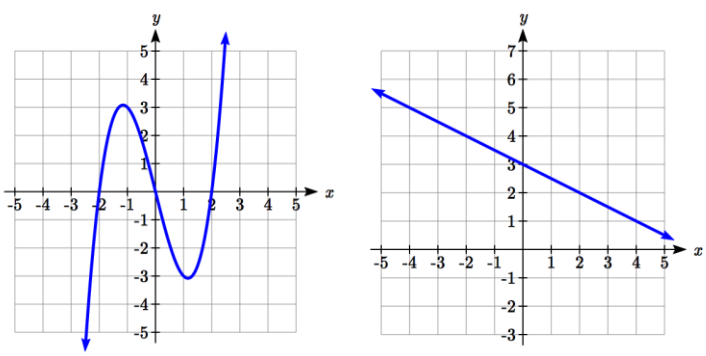


Oftentimes a graph of a relationship can be used to define a function. By convention, graphs are typically created with the input quantity along the horizontal axis and the output quantity along the vertical.

The most common graph has y on the vertical axis and x on the horizontal axis, and we say y is a function of x, or $y = f(x)$ when the function is named f.

Example 8

Which of these graphs defines a function $y=f(x)$? Which of these graphs defines a one-to-one function?



Looking at the three graphs above, the first two define a function $y=f(x)$, since for each input value along the horizontal axis there is exactly one output value corresponding, determined by the y-value of the graph. The 3rd graph does not define a function $y=f(x)$ since some input values, such as $x=2$, correspond with more than one output value.

Graph 1 is not a one-to-one function. For example, the output value 3 has two corresponding input values, -1 and 2.3

Graph 2 is a one-to-one function; each input corresponds to exactly one output, and every output corresponds to exactly one input.

Graph 3 is not even a function so there is no reason to even check to see if it is a one-to-one function.

Vertical Line Test

The **vertical line test** is a handy way to think about whether a graph defines the vertical output as a function of the horizontal input. Imagine drawing vertical lines through the graph. If any vertical line would cross the graph more than once, then the graph does not define only one vertical output for each horizontal input.

Horizontal Line Test

Once you have determined that a graph defines a function, an easy way to determine if it is a one-to-one function is to use the **horizontal line test**. Draw horizontal lines through the graph. If any horizontal line crosses the graph more than once, then the graph does not define a one-to-one function.

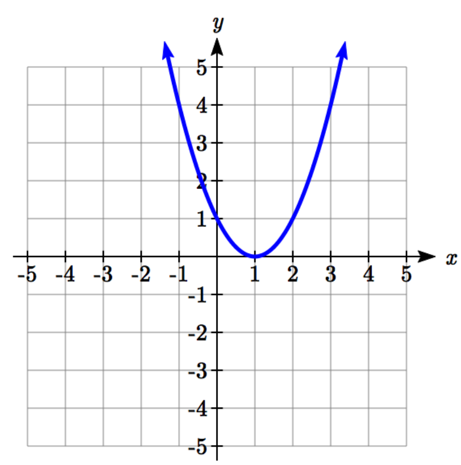
Evaluating a function using a graph requires taking the given input and using the graph to look up the corresponding output. Solving a function equation using a graph requires taking the given output and looking on the graph to determine the corresponding input.

Example 9

Given the graph of $f(x)$

a) Evaluate $f(2)$

b) Solve $f(x) = 4$



a) To evaluate $f(2)$, we find the input of x=2 on the horizontal axis. Moving up to the graph gives the point (2, 1), giving an output of y=1. $f(2) = 1$.

b) To solve $f(x) = 4$, we find the value 4 on the vertical axis because if $f(x) = 4$ then 4 is the output. Moving horizontally across the graph gives two points with the output of 4: (-1,4) and (3,4). These give the two solutions to $f(x) = 4$: $x = -1$ or $x = 3$

This means $f(-1)=4$ and $f(3)=4$, or when the input is -1 or 3, the output is 4.

Notice that while the graph in the previous example is a function, getting two input values for the output value of 4 shows us that this function is not one-to-one.

Try it Now

5. Using the graph from example 9, solve $f(x)=1$.

Formulas as Functions

When possible, it is very convenient to define relationships using formulas. If it is possible to express the output as a formula involving the input quantity, then we can define a function.

Example 10

Express the relationship $2n + 6p = 12$ as a function $p = f(n)$ if possible.

To express the relationship in this form, we need to be able to write the relationship where p is a function of n, which means writing it as p = [something involving n].

$2n + 6p = 12$ subtract $2n$ from both sides
$6p = 12 - 2n$ divide both sides by 6 and simplify

$$p = \frac{12-2n}{6} = \frac{12}{6} - \frac{2n}{6} = 2 - \frac{1}{3}n$$

Having rewritten the formula as p=, we can now express p as a function:

$$p = f(n) = 2 - \frac{1}{3}n$$

It is important to note that not every relationship can be expressed as a function with a formula.

Note the important feature of an equation written as a function is that the output value can be determined directly from the input by doing evaluations - no further solving is required. This allows the relationship to act as a magic box that takes an input, processes it, and returns an output. Modern technology and computers rely on these functional relationships, since the evaluation of the function can be programmed into machines, whereas solving things is much more challenging.

Example 11

Express the relationship $x^2 + y^2 = 1$ as a function $y = f(x)$ if possible.

If we try to solve for y in this equation:

$$y^2 = 1 - x^2$$
$$y = \pm\sqrt{1-x^2}$$

We end up with two outputs corresponding to the same input, so this relationship cannot be represented as a single function $y = f(x)$.

As with tables and graphs, it is common to evaluate and solve functions involving formulas. Evaluating will require replacing the input variable in the formula with the value provided and calculating. Solving will require replacing the output variable in the formula with the value provided, and solving for the input(s) that would produce that output.

Example 12

Given the function $k(t) = t^3 + 2$

a) Evaluate $k(2)$
b) Solve $k(t) = 1$

a) To evaluate $k(2)$, we plug in the input value 2 into the formula wherever we see the input variable t, then simplify

$k(2) = 2^3 + 2$

$k(2) = 8 + 2$

So $k(2) = 10$

b) To solve $k(t) = 1$, we set the formula for $k(t)$ equal to 1, and solve for the input value that will produce that output

$k(t) = 1$	substitute the original formula $k(t) = t^3 + 2$
$t^3 + 2 = 1$	subtract 2 from each side
$t^3 = -1$	take the cube root of each side
$t = -1$	

When solving an equation using formulas, you can check your answer by using your solution in the original equation to see if your calculated answer is correct.

We want to know if $k(t) = 1$ is true when $t = -1$.

$k(-1) = (-1)^3 + 2$

$= -1 + 2$

$= 1$ which was the desired result.

Example 13

Given the function $h(p) = p^2 + 2p$

a) Evaluate $h(4)$
b) Solve $h(p) = 3$

To evaluate $h(4)$ we substitute the value 4 for the input variable p in the given function.

a) $h(4) = (4)^2 + 2(4)$

$= 16 + 8$

$= 24$

b) *h(p)* = 3 Substitute the original function $h(p) = p^2 + 2p$

$p^2 + 2p = 3$ This is quadratic, so we can rearrange the equation to get it = 0

$p^2 + 2p - 3 = 0$ subtract 3 from each side

$p^2 + 2p - 3 = 0$ this is factorable, so we factor it

$(p+3)(p-1) = 0$

By the zero factor theorem since $(p+3)(p-1) = 0$, either $(p+3) = 0$ or $(p-1) = 0$ (or both of them equal 0) and so we solve both equations for p, finding p = -3 from the first equation and p = 1 from the second equation.

This gives us the solution: *h(p)* = 3 when p = 1 or p = -3

We found two solutions in this case, which tells us this function is not one-to-one.

Try it Now

6. Given the function $g(m) = \sqrt{m-4}$
   a. Evaluate $g(5)$
   b. Solve *g(m)* = 2

Basic Toolkit Functions

In this text, we will be exploring functions – the shapes of their graphs, their unique features, their equations, and how to solve problems with them. When learning to read, we start with the alphabet. When learning to do arithmetic, we start with numbers. When working with functions, it is similarly helpful to have a base set of elements to build from. We call these our "toolkit of functions" – a set of basic named functions for which we know the graph, equation, and special features.

For these definitions we will use x as the input variable and *f(x)* as the output variable.

Toolkit Functions	
Linear	
Constant:	$f(x) = c$, where c is a constant (number)
Identity:	$f(x) = x$
Absolute Value:	$f(x) = \lvert x \rvert$
Power	
Quadratic:	$f(x) = x^2$
Cubic:	$f(x) = x^3$
Reciprocal:	$f(x) = \frac{1}{x}$
Reciprocal squared:	$f(x) = \frac{1}{x^2}$
Square root:	$f(x) = \sqrt[2]{x} = \sqrt{x}$
Cube root:	$f(x) = \sqrt[3]{x}$

You will see these toolkit functions, combinations of toolkit functions, their graphs and their transformations frequently throughout this book. In order to successfully follow along later in the book, it will be very helpful if you can recognize these toolkit functions and their features quickly by name, equation, graph and basic table values.

Not every important equation can be written as $y = f(x)$. An example of this is the equation of a circle. Recall the distance formula for the distance between two points:

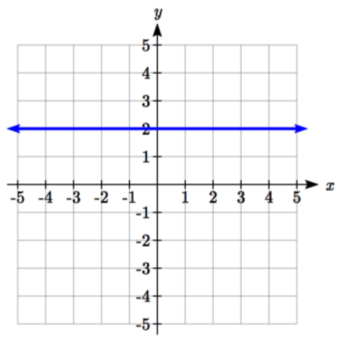
$dist = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$

A circle with radius r with center at (h, k) can be described as all points (x, y) a distance of r from the center, so using the distance formula, $r = \sqrt{(x-h)^2 + (y-k)^2}$, giving

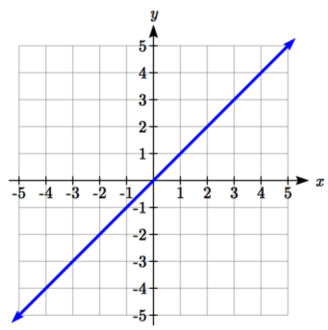
Equation of a circle
A circle with radius r with center (h, k) has equation $r^2 = (x-h)^2 + (y-k)^2$

Graphs of the Toolkit Functions

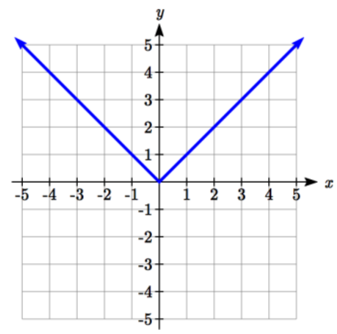
Constant Function: $f(x) = 2$



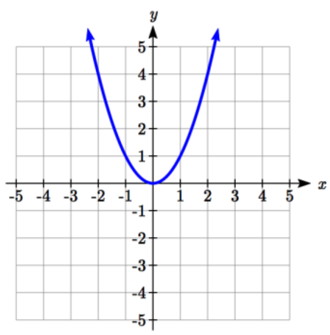
Identity: $f(x) = x$



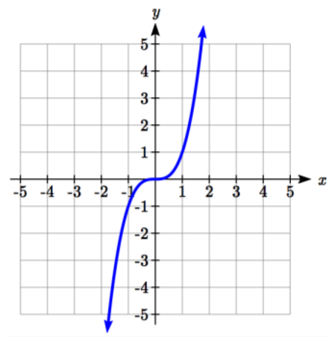
Absolute Value: $f(x) = |x|$



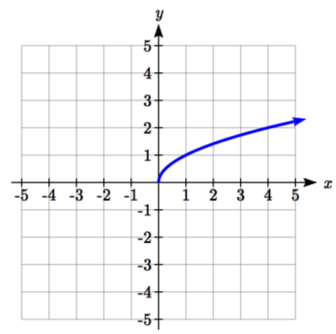
Quadratic: $f(x) = x^2$



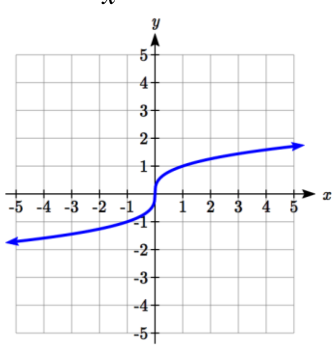
Cubic: $f(x) = x^3$



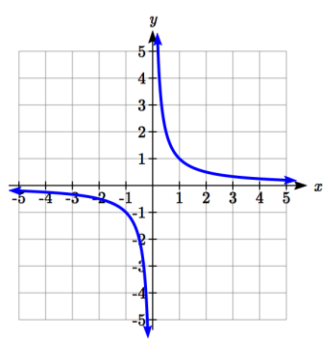
Square root: $f(x) = \sqrt{x}$



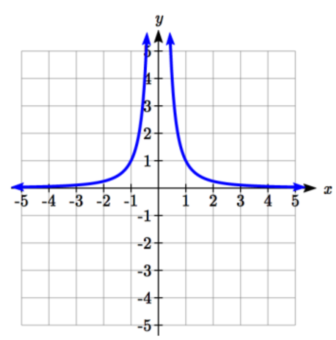
Cube root: $f(x) = \sqrt[3]{x}$



Reciprocal: $f(x) = \dfrac{1}{x}$



Reciprocal squared: $f(x) = \dfrac{1}{x^2}$



Important Topics of this Section
Definition of a function
Input (independent variable)
Output (dependent variable)
Definition of a one-to-one function
Function notation
Descriptive variables
Functions in words, tables, graphs & formulas
Vertical line test
Horizontal line test
Evaluating a function at a specific input value
Solving a function given a specific output value
Toolkit Functions

Try it Now Answers

1. Yes: for each bank account, there would be one balance associated

2. No: there could be several bank accounts with the same balance

3. Yes it's a function; No, it's not one-to-one (several percents give the same letter grade)

4. When n=4, Q=g(4)=6

5. There are two points where the output is 1: $x = 0$ or $x = 2$

6. a. $g(5) = \sqrt{5-4} = 1$

   b. $\sqrt{m-4} = 2$. Square both sides to get $m - 4 = 4$. $m = 8$

Section 1.1 Exercises

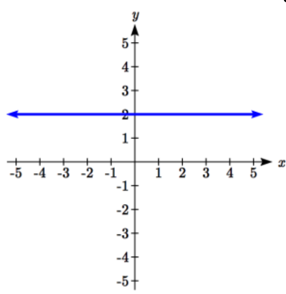
1. The amount of garbage, G, produced by a city with population p is given by $G = f(p)$. G is measured in tons per week, and p is measured in thousands of people.
    a. The town of Tola has a population of 40,000 and produces 13 tons of garbage each week. Express this information in terms of the function f.
    b. Explain the meaning of the statement $f(5) = 2$.

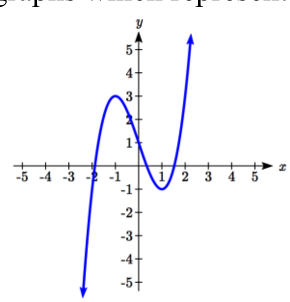
2. The number of cubic yards of dirt, D, needed to cover a garden with area a square feet is given by $D = g(a)$.
    a. A garden with area 5000 ft^2 requires 50 cubic yards of dirt. Express this information in terms of the function g.
    b. Explain the meaning of the statement $g(100) = 1$.

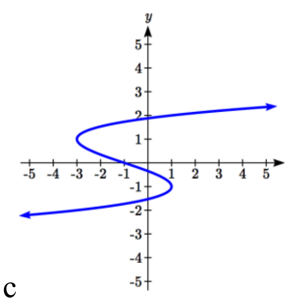
3. Let $f(t)$ be the number of ducks in a lake t years after 1990. Explain the meaning of each statement:
    a. $f(5) = 30$ b. $f(10) = 40$

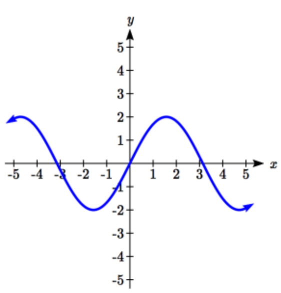
4. Let $h(t)$ be the height above ground, in feet, of a rocket t seconds after launching. Explain the meaning of each statement:
    a. $h(1) = 200$ b. $h(2) = 350$

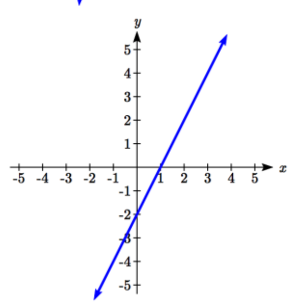
5. Select all of the following graphs which represent y as a function of x.

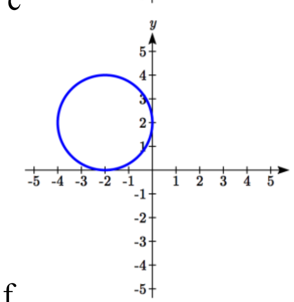

a


b

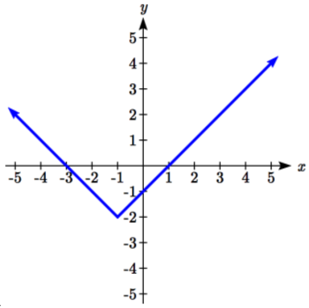

c


d

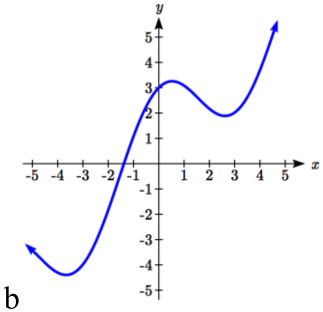

e


f

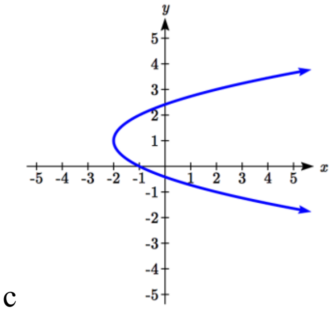
6. Select all of the following graphs which represent y as a function of x.



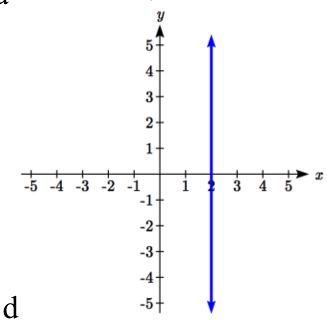
a



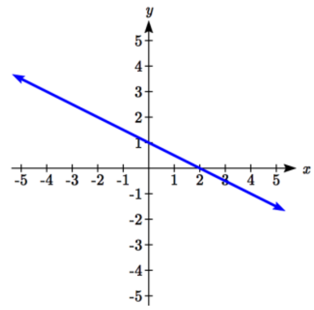
b



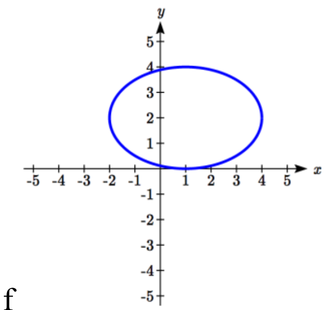
c



d



e



f

7. Select all of the following tables which represent y as a function of x.

a.

x	5	10	15
y	3	8	14

b.

x	5	10	15
y	3	8	8

c.

x	5	10	10
y	3	8	14

8. Select all of the following tables which represent y as a function of x.

a.

x	2	6	13
y	3	10	10

b.

x	2	6	6
y	3	10	14

c.

x	2	6	13
y	3	10	14

9. Select all of the following tables which represent y as a function of x.

a.

x	**y**
0	-2
3	1
4	6
8	9
3	1

b.

x	**y**
-1	-4
2	3
5	4
8	7
12	11

c.

x	**y**
0	-5
3	1
3	4
9	8
16	13

d.

x	**y**
-1	-4
1	2
4	2
9	7
12	13

10. Select all of the following tables which represent y as a function of x.

a.

x	**y**
-4	-2
3	2
6	4
9	7
12	16

b.

x	**y**
-5	-3
2	1
2	4
7	9
11	10

c.

x	**y**
-1	-3
1	2
5	4
9	8
1	2

d.

x	**y**
-1	-5
3	1
5	1
8	7
14	12

11. Select all of the following tables which represent y as a function of x **and** are one-to-one.

a.

x	3	8	12
y	4	7	7

b.

x	3	8	12
y	4	7	13

c.

x	3	8	8
y	4	7	13

12. Select all of the following tables which represent y as a function of x **and** are one-to-one.

a.

x	2	8	8
y	5	6	13

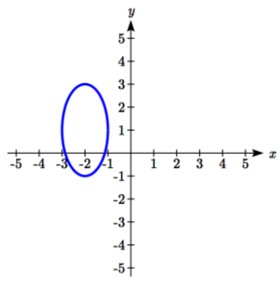
b.

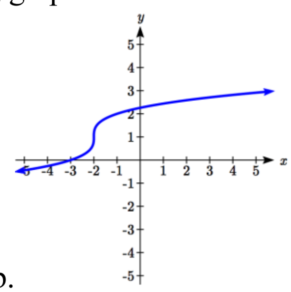
x	2	8	14
y	5	6	6

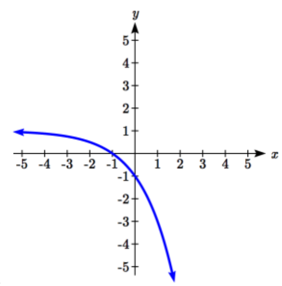
c.

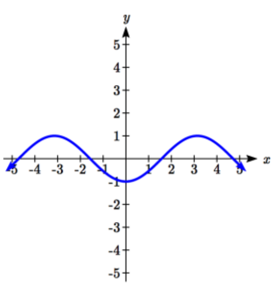
x	2	8	14
y	5	6	13

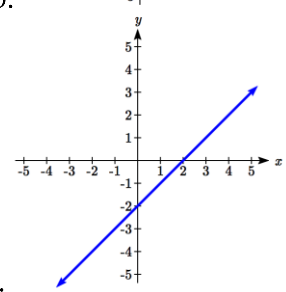
13. Select all of the following graphs which are **one-to-one functions**.

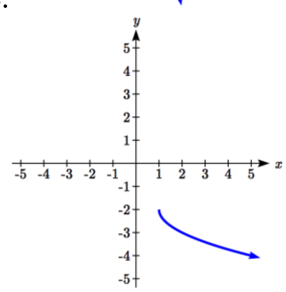

a.


b.

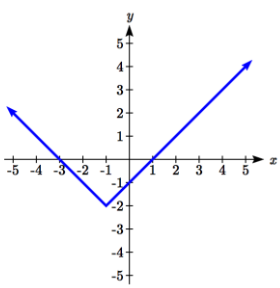

c.

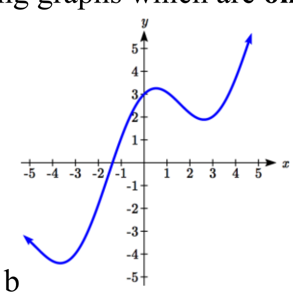

d.

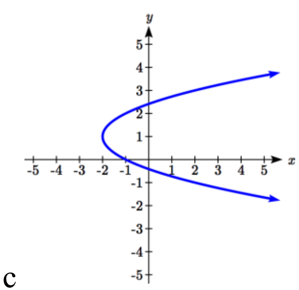

e.

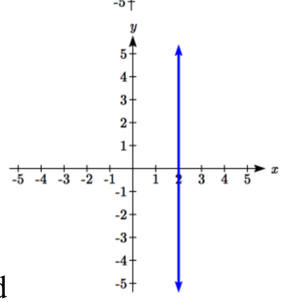

f.

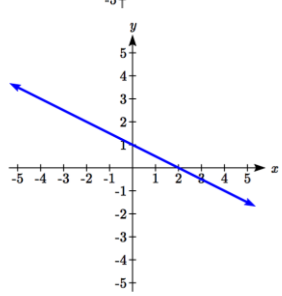
14. Select all of the following graphs which are **one-to-one functions**.

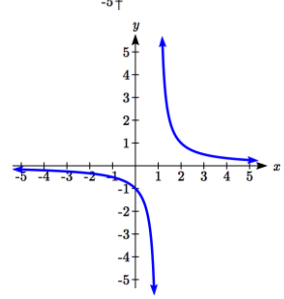

a


b

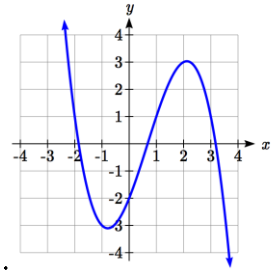

c

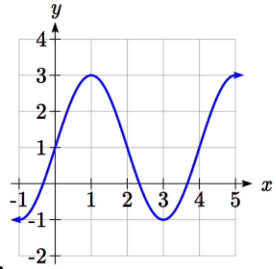

d


e


f

Given each function $f(x)$ graphed, evaluate $f(1)$ and $f(3)$

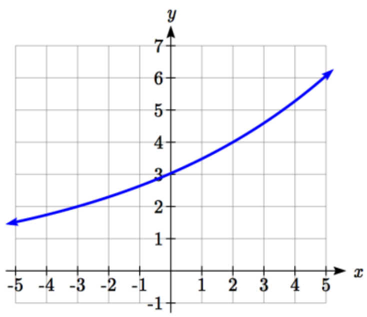
15. 

16. 

17. Given the function $g(x)$ graphed here,

a. Evaluate $g(2)$

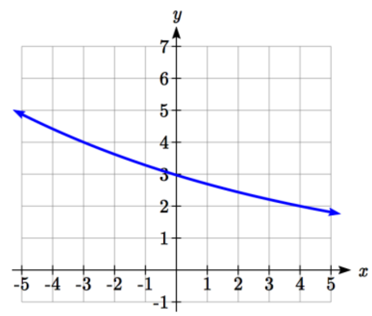
b. Solve $g(x) = 2$



18. Given the function $f(x)$ graphed here.

a. Evaluate $f(4)$

b. Solve $f(x)=4$



19. Based on the table below,

a. Evaluate $f(3)$ b. Solve $f(x)=1$

$\boldsymbol{x}$	0	1	2	3	4	5	6	7	8	9
$\boldsymbol{f(x)}$	74	28	1	53	56	3	36	45	14	47

20. Based on the table below,

a. Evaluate $f(8)$ b. Solve $f(x)=7$

$\boldsymbol{x}$	0	1	2	3	4	5	6	7	8	9
$\boldsymbol{f(x)}$	62	8	7	38	86	73	70	39	75	34

For each of the following functions, evaluate: $f(-2)$, $f(-1)$, $f(0)$, $f(1)$, and $f(2)$

21. $f(x) = 4 - 2x$

22. $f(x) = 8 - 3x$

23. $f(x) = 8x^2 - 7x + 3$

24. $f(x) = 6x^2 - 7x + 4$

25. $f(x) = -x^3 + 2x$

26. $f(x) = 5x^4 + x^2$

27. $f(x) = 3 + \sqrt{x+3}$

28. $f(x) = 4 - \sqrt[3]{x-2}$

29. $f(x) = (x-2)(x+3)$

30. $f(x) = (x+3)(x-1)^2$

31. $f(x) = \dfrac{x-3}{x+1}$

32. $f(x) = \dfrac{x-2}{x+2}$

33. $f(x) = 2^x$

34. $f(x) = 3^x$

35. Suppose $f(x) = x^2 + 8x - 4$. Compute the following:

a. $f(-1) + f(1)$ b. $f(-1) - f(1)$

36. Suppose $f(x) = x^2 + x + 3$. Compute the following:

a. $f(-2) + f(4)$ b. $f(-2) - f(4)$

37. Let $f(t) = 3t + 5$

a. Evaluate $f(0)$ b. Solve $f(t) = 0$

38. Let $g(p) = 6 - 2p$

a. Evaluate $g(0)$ b. Solve $g(p) = 0$

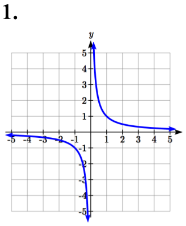
39. Match each function name with its equation.

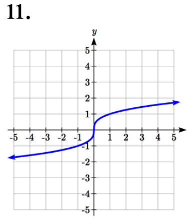
a. $y = x$
b. $y = x^3$
c. $y = \sqrt[3]{x}$
d. $y = \dfrac{1}{x}$
e. $y = x^2$
f. $y = \sqrt{x}$
g. $y = |x|$
h. $y = \dfrac{1}{x^2}$

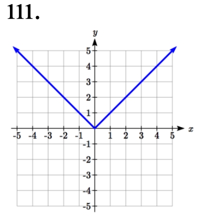
i. Cube root
ii. Reciprocal
iii. Linear
iv. Square Root
v. Absolute Value
vi. Quadratic
vii. Reciprocal Squared
viii. Cubic

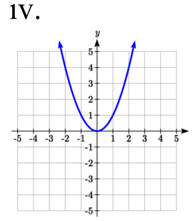
40. Match each graph with its equation.

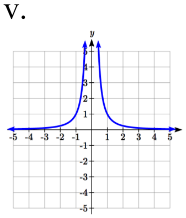
a. $y = x$
b. $y = x^3$
c. $y = \sqrt[3]{x}$
d. $y = \dfrac{1}{x}$
e. $y = x^2$
f. $y = \sqrt{x}$
g. $y = |x|$
h. $y = \dfrac{1}{x^2}$

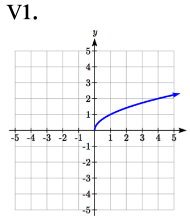
i.


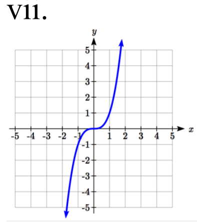
ii.


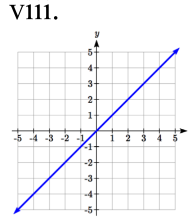
iii.


iv.


v.


vi.


vii.


viii.


41. Match each table with its equation.

a. $y = x^2$
b. $y = x$
c. $y = \sqrt{x}$
d. $y = 1/x$
e. $y = |x|$
f. $y = x^3$

i.

In	Out
-2	-0.5
-1	-1
0	_
1	1
2	0.5
3	0.33

ii.

In	Out
-2	-2
-1	-1
0	0
1	1
2	2
3	3

iii.

In	Out
-2	-8
-1	-1
0	0
1	1
2	8
3	27

iv.

In	Out
-2	4
-1	1
0	0
1	1
2	4
3	9

v.

In	Out
-2	_
-1	_
0	0
1	1
4	2
9	3

vi.

In	Out
-2	2
-1	1
0	0
1	1
2	2
3	3

42. Match each equation with its table
    a. Quadratic
    b. Absolute Value
    c. Square Root
    d. Linear
    e. Cubic
    f. Reciprocal

i.

In	Out
-2	-0.5
-1	-1
0	_
1	1
2	0.5
3	0.33

ii.

In	Out
-2	-2
-1	-1
0	0
1	1
2	2
3	3

iii.

In	Out
-2	-8
-1	-1
0	0
1	1
2	8
3	27

iv.

In	Out
-2	4
-1	1
0	0
1	1
2	4
3	9

v.

In	Out
-2	_
-1	_
0	0
1	1
4	2
9	3

vi.

In	Out
-2	2
-1	1
0	0
1	1
2	2
3	3

43. Write the equation of the circle centered at $(3, -9)$ with radius 6.

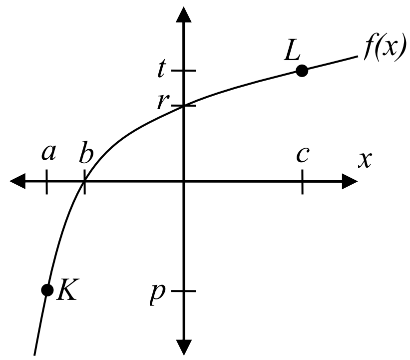
44. Write the equation of the circle centered at $(9, -8)$ with radius 11.

45. Sketch a reasonable graph for each of the following functions. [UW]
    a. Height of a person depending on age.
    b. Height of the top of your head as you jump on a pogo stick for 5 seconds.
    c. The amount of postage you must put on a first class letter, depending on the weight of the letter.

46. Sketch a reasonable graph for each of the following functions. [UW]

a. Distance of your big toe from the ground as you ride your bike for 10 seconds.
b. Your height above the water level in a swimming pool after you dive off the high board.
c. The percentage of dates and names you'll remember for a history test, depending on the time you study.

47. Using the graph shown,
 a. Evaluate $f(c)$
 b. Solve $f(x) = p$
 c. Suppose $f(b) = z$. Find $f(z)$
 d. What are the coordinates of points *L* and *K*?



48. Dave leaves his office in Padelford Hall on his way to teach in Gould Hall. Below are several different scenarios. In each case, sketch a plausible (reasonable) graph of the function $s = d(t)$ which keeps track of Dave's distance s from Padelford Hall at time t. Take distance units to be "feet" and time units to be "minutes." Assume Dave's path to Gould Hall is long a straight line which is 2400 feet long. [UW]

 a. Dave leaves Padelford Hall and walks at a constant spend until he reaches Gould Hall 10 minutes later.

 b. Dave leaves Padelford Hall and walks at a constant speed. It takes him 6 minutes to reach the half-way point. Then he gets confused and stops for 1 minute. He then continues on to Gould Hall at the same constant speed he had when he originally left Padelford Hall.

 c. Dave leaves Padelford Hall and walks at a constant speed. It takes him 6 minutes to reach the half-way point. Then he gets confused and stops for 1 minute to figure out where he is. Dave then continues on to Gould Hall at twice the constant speed he had when he originally left Padelford Hall.

d. Dave leaves Padelford Hall and walks at a constant speed. It takes him 6 minutes to reach the half-way point. Then he gets confused and stops for 1 minute to figure out where he is. Dave is totally lost, so he simply heads back to his office, walking the same constant speed he had when he originally left Padelford Hall.

e. Dave leaves Padelford heading for Gould Hall at the same instant Angela leaves Gould Hall heading for Padelford Hall. Both walk at a constant speed, but Angela walks twice as fast as Dave. Indicate a plot of "distance from Padelford" vs. "time" for the both Angela and Dave.

f. Suppose you want to sketch the graph of a new function $s = g(t)$ that keeps track of Dave's distance s from Gould Hall at time t. How would your graphs change in (a)-(e)?

Chapter 2: Linear Functions

Chapter one was a window that gave us a peek into the entire course. Our goal was to understand the basic structure of functions and function notation, the toolkit functions, domain and range, how to recognize and understand composition and transformations of functions and how to understand and utilize inverse functions. With these basic components in hand we will further research the specific details and intricacies of each type of function in our toolkit and use them to model the world around us.

Mathematical Modeling
As we approach day to day life we often need to quantify the things around us, giving structure and numeric value to various situations. This ability to add structure enables us to make choices based on patterns we see that are weighted and systematic. With this structure in place we can model and even predict behavior to make decisions. Adding a numerical structure to a real world situation is called **Mathematical Modeling.**

When modeling real world scenarios, there are some common growth patterns that are regularly observed. We will devote this chapter and the rest of the book to the study of the functions used to model these growth patterns.

Section 2.1 Linear Functions

As you hop into a taxicab in Las Vegas, the meter will immediately read \$3.50; this is the "drop" charge made when the taximeter is activated. After that initial fee, the taximeter will add \$2.76 for each mile the taxi drives[1]. In this scenario, the total taxi fare depends upon the number of miles ridden in the taxi, and we can ask whether it is possible to model this type of scenario with a function. Using descriptive variables, we choose m for miles and C for Cost in dollars as a function of miles: $C(m)$.

[1] Nevada Taxicab Authority, retrieved Aug 4, 2020. There is also a waiting fee assessed when the taxi is waiting at red lights, but we'll ignore that in this discussion.

We know for certain that $C(0) = 3.50$, since the \$3.50 drop charge is assessed regardless of how many miles are driven. Since \$2.67 is added for each mile driven, then $C(1) = 3.50 + 2.67 = 6.17$.

If we then drove a second mile, another \$2.67 would be added to the cost:
$C(2) = 3.50 + 2.67 + 2.67 = 3.50 + 2.67(2) = 8.84$

If we drove a third mile, another \$2.67 would be added to the cost:
$C(3) = 3.50 + 2.67 + 2.67 + 2.67 = 3.50 + 2.67(3) = 11.51$

From this we might observe the pattern, and conclude that if m miles are driven, $C(m) = 3.50 + 2.67m$ because we start with a \$3.50 drop fee and then for each mile increase we add \$2.67.

It is good to verify that the units make sense in this equation. The \$3.50 drop charge is measured in dollars; the \$2.67 charge is measured in dollars per mile.

$$C(m) = 3.50\,dollars + \left(2.67\,\frac{dollars}{mile}\right)(m\ miles)$$

When dollars per mile are multiplied by a number of miles, the result is a number of dollars, matching the units on the 3.50, and matching the desired units for the *C* function.

Notice this equation $C(m) = 3.50 + 2.67m$ consisted of two quantities. The first is the fixed \$3.50 charge which does not change based on the value of the input. The second is the \$2.67 dollars per mile value, which is a **rate of change**. In the equation, this rate of change is multiplied by the input value.

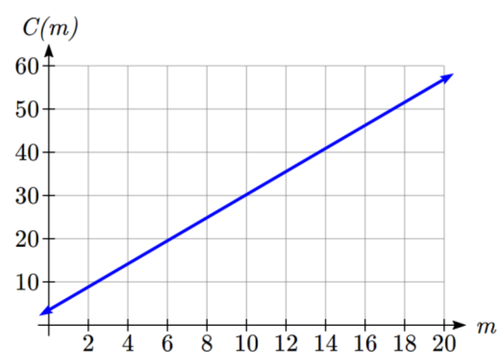
Looking at this same problem in table format we can also see the cost changes by \$2.67 for every 1 mile increase.

m	0	1	2	3
$C(m)$	3.50	6.17	8.84	11.51

It is important here to note that in this equation, the **rate of change is constant**; over any interval, the rate of change is the same.

Graphing this equation, $C(m) = 3.50 + 2.67m$ we see the shape is a line, which is how these functions get their name: **linear functions**.

When the number of miles is zero the cost is \$3.50, giving the point (0, 3.50) on the graph. This is the vertical or *C(m)* intercept. The graph is increasing in a straight line from left to right because for each mile the cost goes up by \$2.67; this rate remains consistent.



In this example, you have seen the taxicab cost modeled in words, an equation, a table and in graphical form. Whenever possible, ensure that you can link these four representations together to continually build your skills. It is important to note that you will not always be able to find all 4 representations for a problem and so being able to work with all 4 forms is very important.

Linear Function

A **linear function** is a function whose graph produces a line. Linear functions can always be written in the form

$f(x) = b + mx$ or $f(x) = mx + b$; they're equivalent

where

b is the initial or starting value of the function (when input, $x = 0$), and

m is the constant rate of change of the function

Many people like to write linear functions in the form $f(x) = b + mx$ because it corresponds to the way we tend to speak: "The output starts at b and increases at a rate of m."

For this reason alone we will use the $f(x) = b + mx$ form for many of the examples, but remember they are equivalent and can be written correctly both ways.

Slope and Increasing/Decreasing

m is the constant rate of change of the function (also called **slope**). The slope determines if the function is an increasing function or a decreasing function.

$f(x) = b + mx$ is an **increasing** function if $m > 0$

$f(x) = b + mx$ is a **decreasing** function if $m < 0$

If $m = 0$, the rate of change zero, and the function $f(x) = b + 0x = b$ is just a horizontal line passing through the point (0, b), neither increasing nor decreasing.

Example 1

Marcus currently owns 200 songs in his iTunes collection. Every month, he adds 15 new songs. Write a formula for the number of songs, N, in his iTunes collection as a function of the number of months, m. How many songs will he own in a year?

The initial value for this function is 200, since he currently owns 200 songs, so $N(0) = 200$. The number of songs increases by 15 songs per month, so the rate of change is 15 songs per month. With this information, we can write the formula: $N(m) = 200 + 15m$.

N(m) is an increasing linear function. With this formula we can predict how many songs he will have in 1 year (12 months):
$N(12) = 200 + 15(12) = 200 + 180 = 380$. Marcus will have 380 songs in 12 months.

Try it Now

1. If you earn $30,000 per year and you spend $29,000 per year write an equation for the amount of money you save after *y* years, if you start with nothing.
"The most important thing, spend less than you earn![2]"

Calculating Rate of Change

Given two values for the input, x_1 and x_2, and two corresponding values for the output, y_1 and y_2, or a set of points, $(x_1,\ y_1)$ and $(x_2,\ y_2)$, if we wish to find a linear function that contains both points we can calculate the rate of change, *m*:

$$m = \frac{\text{change in output}}{\text{change in input}} = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

Rate of change of a linear function is also called the **slope** of the line.

Note in function notation, $y_1 = f(x_1)$ and $y_2 = f(x_2)$, so we could equivalently write

$$m = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

Example 2

The population of a city increased from 23,400 to 27,800 between 2002 and 2006. Find the rate of change of the population during this time span.

The rate of change will relate the change in population to the change in time. The population increased by $27800 - 23400 = 4400$ people over the 4 year time interval. To find the rate of change, the number of people per year the population changed by:

$$\frac{4400\ people}{4\ years} = 1100\ \frac{people}{year} = 1100 \text{ people per year}$$

Notice that we knew the population was increasing, so we would expect our value for *m* to be positive. This is a quick way to check to see if your value is reasonable.

[2] http://www.thesimpledollar.com/2009/06/19/rule-1-spend-less-than-you-earn/

Example 3

The pressure, P, in pounds per square inch (PSI) on a diver depends upon their depth below the water surface, d, in feet, following the equation $P(d) = 14.696 + 0.434d$. Interpret the components of this function.

The rate of change, or slope, 0.434 would have units $\frac{\text{output}}{\text{input}} = \frac{\text{pressure}}{\text{depth}} = \frac{\text{PSI}}{\text{ft}}$. This tells us the pressure on the diver increases by 0.434 PSI for each foot their depth increases.

The initial value, 14.696, will have the same units as the output, so this tells us that at a depth of 0 feet, the pressure on the diver will be 14.696 PSI.

Example 4

If $f(x)$ is a linear function, $f(3) = -2$, and $f(8) = 1$, find the rate of change.

$f(3) = -2$ tells us that the input 3 corresponds with the output -2, and $f(8) = 1$ tells us that the input 8 corresponds with the output 1. To find the rate of change, we divide the change in output by the change in input:

$m = \frac{\text{change in output}}{\text{change in input}} = \frac{1-(-2)}{8-3} = \frac{3}{5}$. If desired we could also write this as $m = 0.6$

Note that it is not important which pair of values comes first in the subtractions so long as the first output value used corresponds with the first input value used.

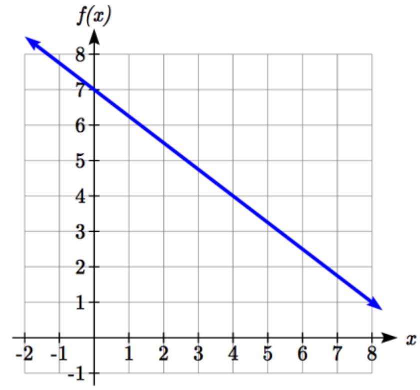
Try it Now

2. Given the two points (2, 3) and (0, 4), find the rate of change. Is this function increasing or decreasing?

We can now find the rate of change given two input-output pairs, and can write an equation for a linear function once we have the rate of change and initial value. If we have two input-output pairs and they do not include the initial value of the function, then we will have to solve for it.

Example 5

Write an equation for the linear function graphed to the right.



Looking at the graph, we might notice that it passes through the points (0, 7) and (4, 4). From the first value, we know the initial value of the function is $b = 7$, so in this case we will only need to calculate the rate of change:

$$m = \frac{4-7}{4-0} = \frac{-3}{4}$$

This allows us to write the equation:

$$f(x) = 7 - \frac{3}{4}x$$

Example 6

If $f(x)$ is a linear function, $f(3) = -2$, and $f(8) = 1$, find an equation for the function.

In example 3, we computed the rate of change to be $m = \frac{3}{5}$. In this case, we do not know the initial value $f(0)$, so we will have to solve for it. Using the rate of change, we know the equation will have the form $f(x) = b + \frac{3}{5}x$. Since we know the value of the function when $x = 3$, we can evaluate the function at 3.

$f(3) = b + \frac{3}{5}(3)$ Since we know that $f(3) = -2$, we can substitute on the left side

$-2 = b + \frac{3}{5}(3)$ This leaves us with an equation we can solve for the initial value

$b = -2 - \frac{9}{5} = \frac{-19}{5}$

Combining this with the value for the rate of change, we can now write a formula for this function:

$$f(x) = \frac{-19}{5} + \frac{3}{5}x$$

Example 7

Working as an insurance salesperson, Ilya earns a base salary and a commission on each new policy, so Ilya's weekly income, I, depends on the number of new policies, n, he sells during the week. Last week he sold 3 new policies, and earned \$760 for the week. The week before, he sold 5 new policies, and earned \$920. Find an equation for *I(n)*, and interpret the meaning of the components of the equation.

The given information gives us two input-output pairs: (3,760) and (5,920). We start by finding the rate of change.

$$m = \frac{920-760}{5-3} = \frac{160}{2} = 80$$

Keeping track of units can help us interpret this quantity. Income increased by \$160 when the number of policies increased by 2, so the rate of change is \$80 per policy; Ilya earns a commission of \$80 for each policy sold during the week.

We can then solve for the initial value

$I(n) = b + 80n$ then when $n = 3$, $I(3) = 760$, giving

$760 = b + 80(3)$ this allows us to solve for b

$b = 760 - 80(3) = 520$

This value is the starting value for the function. This is Ilya's income when $n = 0$, which means no new policies are sold. We can interpret this as Ilya's base salary for the week, which does not depend upon the number of policies sold.

Writing the final equation:

$I(n) = 520 + 80n$

Our final interpretation is: Ilya's base salary is \$520 per week and he earns an additional \$80 commission for each policy sold each week.

Flashback

Looking at Example 7:

Determine the independent and dependent variables.
What is a reasonable domain and range?
Is this function one-to-one?

Try it Now

3. The balance in your college payment account, C, is a function of the number of quarters, q, you attend. Interpret the function $C(a) = 20000 - 4000q$ in words. How many quarters of college can you pay for until this account is empty?

Example 8

Given the table below write a linear equation that represents the table values

w, number of weeks	0	2	4	6
$P(w)$, number of rats	1000	1080	1160	1240

We can see from the table that the initial value of rats is 1000 so in the linear format $P(w) = b + mw$, $b = 1000$.

Rather than solving for m, we can notice from the table that the population goes up by 80 for every 2 weeks that pass. This rate is consistent from week 0, to week 2, 4, and 6. The rate of change is 80 rats per 2 weeks. This can be simplified to 40 rats per week and we can write
$P(w) = b + mw$ as $P(w) = 1000 + 40w$

If you didn't notice this from the table you could still solve for the slope using any two points from the table. For example, using (2, 1080) and (6, 1240),

$$m = \frac{1240 - 1080}{6 - 2} = \frac{160}{4} = 40 \text{ rats per week}$$

Important Topics of this Section
Definition of Modeling
Definition of a linear function
Structure of a linear function
Increasing & Decreasing functions
Finding the vertical intercept (0, b)
Finding the slope/rate of change, m
Interpreting linear functions

Try it Now Answers

1. $S(y) = 30{,}000y - 29{,}000y = 1000y$ $1000 is saved each year.

2. $m = \frac{4-3}{0-2} = \frac{1}{-2} = -\frac{1}{2}$; Decreasing because $m < 0$

3. Your College account starts with $20,000 in it and you withdraw $4,000 each quarter (or your account contains $20,000 and decreases by $4000 each quarter.)
   Solving $C(a) = 0$ gives $a = 5$. You can pay for 5 quarters before the money in this account is gone.

Flashback Answers

n (number of policies sold) is the independent variable
I(n) (weekly income as a function of policies sold) is the dependent variable.

A reasonable domain is (0, 15)*
A reasonable range is ($540, $1740)*
*answers may vary given reasoning is stated; 15 is an arbitrary upper limit based on selling 3 policies per day in a 5 day work week and $1740 corresponds with the domain.

Yes this function is one-to-one

Section 2.1 Exercises

1. A town's population has been growing linearly. In 2003, the population was 45,000, and the population has been growing by 1700 people each year. Write an equation, $P(t)$, for the population t years after 2003.

2. A town's population has been growing linearly. In 2005, the population was 69,000, and the population has been growing by 2500 people each year. Write an equation, $P(t)$, for the population t years after 2005.

3. Sonya is currently 10 miles from home, and is walking further away at 2 miles per hour. Write an equation for her distance from home t hours from now.

4. A boat is 100 miles away from the marina, sailing directly towards it at 10 miles per hour. Write an equation for the distance of the boat from the marina after t hours.

5. Timmy goes to the fair with \$40. Each ride costs \$2. How much money will he have left after riding n rides?

6. At noon, a barista notices she has \$20 in her tip jar. If she makes an average of \$0.50 from each customer, how much will she have in her tip jar if she serves n more customers during her shift?

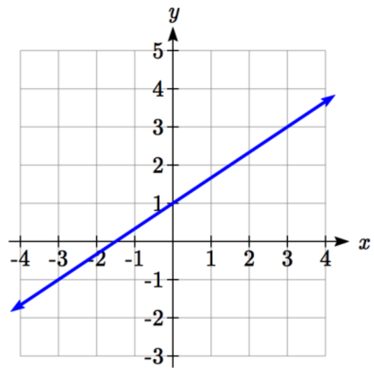
Determine if each function is increasing or decreasing

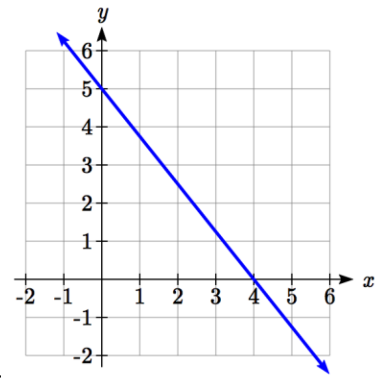
7. $f(x)=4x+3$
8. $g(x)=5x+6$
9. $a(x)=5-2x$
10. $b(x)=8-3x$
11. $h(x)=-2x+4$
12. $k(x)=-4x+1$
13. $j(x)=\frac{1}{2}x-3$
14. $p(x)=\frac{1}{4}x-5$
15. $n(x)=-\frac{1}{3}x-2$
16. $m(x)=-\frac{3}{8}x+3$

Find the slope of the line that passes through the two given points

17. (2, 4) and (4, 10)
18. (1, 5) and (4, 11)
19. (-1,4) and (5, 2)
20. (-2, 8) and (4, 6)
21. (6,11) and (-4,3)
22. (9,10) and (-6,-12)

Find the slope of the lines graphed

23. 

24. 

25. Sonya is walking home from a friend's house. After 2 minutes she is 1.4 miles from home. Twelve minutes after leaving, she is 0.9 miles from home. What is her rate?

26. A gym membership with two personal training sessions costs \$125, while gym membership with 5 personal training sessions costs \$260. What is the rate for personal training sessions?

27. A city's population in the year 1960 was 287,500. In 1989 the population was 275,900. Compute the slope of the population growth (or decline) and make a statement about the population rate of change in people per year.

28. A city's population in the year 1958 was 2,113,000. In 1991 the population was 2,099,800. Compute the slope of the population growth (or decline) and make a statement about the population rate of change in people per year.

29. A phone company charges for service according to the formula: $C(n) = 24 + 0.1n$, where n is the number of minutes talked, and $C(n)$ is the monthly charge, in dollars. Find and interpret the rate of change and initial value.

30. A phone company charges for service according to the formula: $C(n) = 26 + 0.04n$, where n is the number of minutes talked, and $C(n)$ is the monthly charge, in dollars. Find and interpret the rate of change and initial value.

31. Terry is skiing down a steep hill. Terry's elevation, $E(t)$, in feet after t seconds is given by $E(t) = 3000 - 70t$. Write a complete sentence describing Terry's starting elevation and how it is changing over time.

32. Maria is climbing a mountain. Maria's elevation, $E(t)$, in feet after t minutes is given by $E(t) = 1200 + 40t$. Write a complete sentence describing Maria's starting elevation and how it is changing over time.

Given each set of information, find a linear equation satisfying the conditions, if possible

33. $f(-5) = -4$, and $f(5) = 2$

34. $f(-1) = 4$, and $f(5) = 1$

35. Passes through (2, 4) and (4, 10)

36. Passes through (1, 5) and (4, 11)

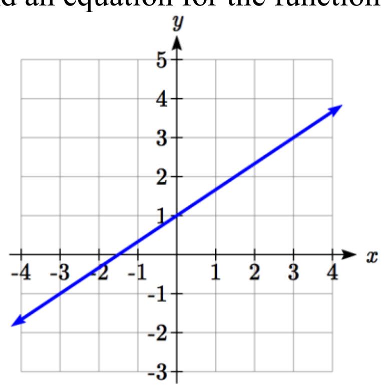
37. Passes through (-1,4) and (5, 2)

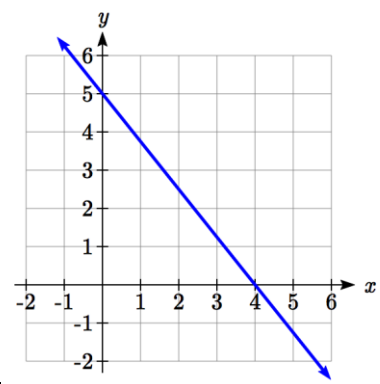
38. Passes through (-2, 8) and (4, 6)

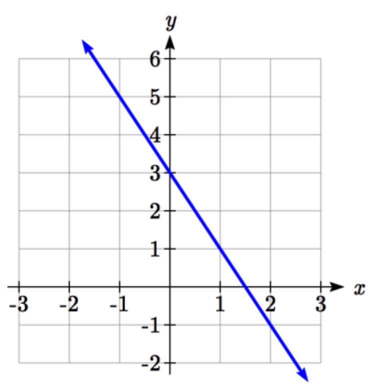
39. x intercept at (-2, 0) and y intercept at (0, -3)

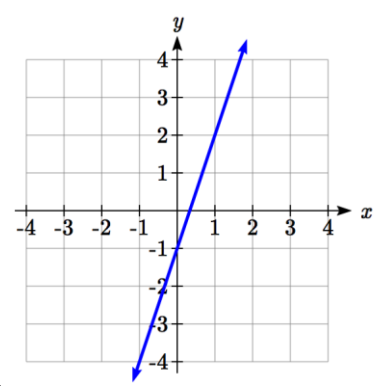
40. x intercept at (-5, 0) and y intercept at (0, 4)

Find an equation for the function graphed

41.


42.


43.


44.


45. A clothing business finds there is a linear relationship between the number of shirts, n, it can sell and the price, p, it can charge per shirt. In particular, historical data shows that 1000 shirts can be sold at a price of $\$30$, while 3000 shirts can be sold at a price of $\$22$. Find a linear equation in the form $p = mn + b$ that gives the price p they can charge for n shirts.

46. A farmer finds there is a linear relationship between the number of bean stalks, n, she plants and the yield, y, each plant produces. When she plants 30 stalks, each plant yields 30 oz of beans. When she plants 34 stalks, each plant produces 28 oz of beans. Find a linear relationships in the form $y = mn + b$ that gives the yield when n stalks are planted.

47. Which of the following tables could represent a linear function? For each that could be linear, find a linear equation models the data.

x	*g(x)*
0	5
5	-10
10	-25
15	-40

x	*h(x)*
0	5
5	30
10	105
15	230

x	*f(x)*
0	-5
5	20
10	45
15	70

x	*k(x)*
5	13
10	28
20	58
25	73

48. Which of the following tables could represent a linear function? For each that could be linear, find a linear equation models the data.

x	*g(x)*
0	6
2	-19
4	-44
6	-69

x	*h(x)*
2	13
4	23
8	43
10	53

x	*f(x)*
2	-4
4	16
6	36
8	56

x	*k(x)*
0	6
2	31
6	106
8	231

49. While speaking on the phone to a friend in Oslo, Norway, you learned that the current temperature there was -23 Celsius (-23°C). After the phone conversation, you wanted to convert this temperature to Fahrenheit degrees, °F, but you could not find a reference with the correct formulas. You then remembered that the relationship between °F and °C is linear. [UW]
   a. Using this and the knowledge that 32°F = 0 °C and 212 °F = 100 °C, find an equation that computes Celsius temperature in terms of Fahrenheit; i.e. an equation of the form C = "an expression involving only the variable F."
   b. Likewise, find an equation that computes Fahrenheit temperature in terms of Celsius temperature; i.e. an equation of the form F = "an expression involving only the variable C."
   c. How cold was it in Oslo in °F?

Section 2.2 Graphs of Linear Functions

When we are working with a new function, it is useful to know as much as we can about the function: its graph, where the function is zero, and any other special behaviors of the function. We will begin this exploration of linear functions with a look at graphs.

When graphing a linear function, there are three basic ways to graph it:

1) By plotting points (at least 2) and drawing a line through the points
2) Using the initial value (output when $x = 0$) and rate of change (slope)
3) Using transformations of the identity function $f(x) = x$

Example 1

Graph $f(x) = 5 - \frac{2}{3}x$ by plotting points

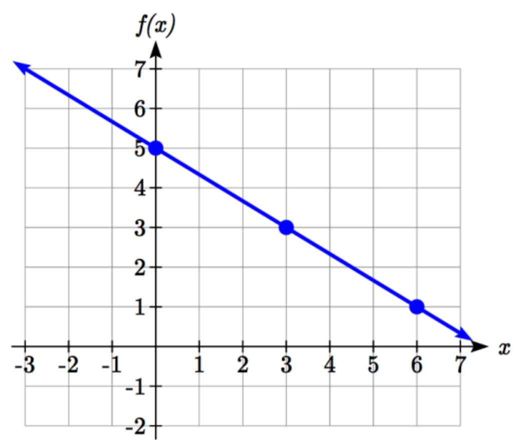
In general, we evaluate the function at two or more inputs to find at least two points on the graph. Usually it is best to pick input values that will "work nicely" in the equation. In this equation, multiples of 3 will work nicely due to the $\frac{2}{3}$ in the equation, and of course using $x = 0$ to get the vertical intercept. Evaluating *f(x)* at $x = 0$, 3 and 6:

$$f(0) = 5 - \frac{2}{3}(0) = 5$$

$$f(3) = 5 - \frac{2}{3}(3) = 3$$

$$f(6) = 5 - \frac{2}{3}(6) = 1$$

These evaluations tell us that the points (0,5), (3,3), and (6,1) lie on the graph of the line. Plotting these points and drawing a line through them gives us the graph.

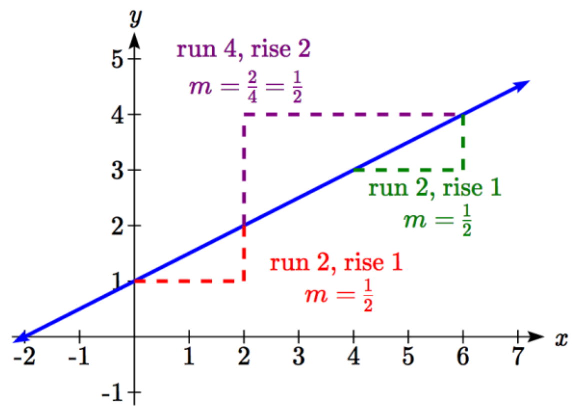


When using the initial value and rate of change to graph, we need to consider the graphical interpretation of these values. Remember the initial value of the function is the output when the input is zero, so in the equation $f(x) = b + mx$, the graph includes the point (0, b). On the graph, this is the vertical intercept – the point where the graph crosses the vertical axis.

For the rate of change, it is helpful to recall that we calculated this value as

$$m = \frac{\text{change of output}}{\text{change of input}}$$

From a graph of a line, this tells us that if we divide the vertical difference, or rise, of the function outputs by the horizontal difference, or run, of the inputs, we will obtain the rate of change, also called slope of the line.



$$m = \frac{\text{change of output}}{\text{change of input}} = \frac{rise}{run}$$

Notice that this ratio is the same regardless of which two points we use.

Graphical Interpretation of a Linear Equation

Graphically, in the equation $f(x) = b + mx$,

b is the **vertical intercept** of the graph and tells us we can start our graph at (0, b)

m is the **slope of the line** and tells us how far to rise & run to get to the next point

Once we have at least 2 points, we can extend the graph of the line to the left and right.

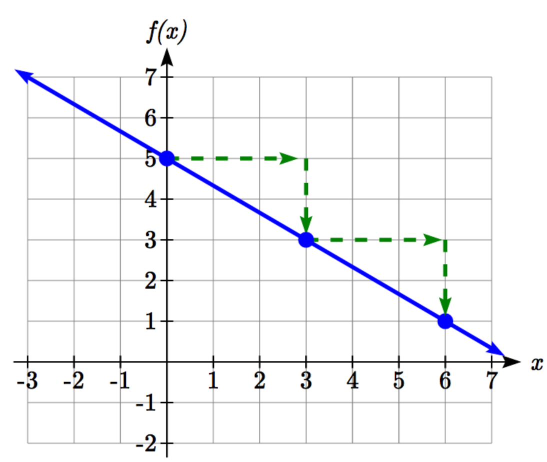
Example 2

Graph $f(x) = 5 - \frac{2}{3}x$ using the vertical intercept and slope.

The vertical intercept of the function is (0, 5), giving us a point on the graph of the line.

The slope is $-\frac{2}{3}$. This tells us that for every 3 units the graph "runs" in the horizontal, the vertical "rise" decreases by 2 units.

In graphing, we can use this by first plotting our vertical intercept on the graph, then using the slope to find a second point. From the initial value (0, 5) the slope tells us that if we move to the right 3, we will move down 2, moving us to the point (3, 3). We can continue this again to find a third point at (6, 1). Finally, extend the line to the left and right, containing these points.

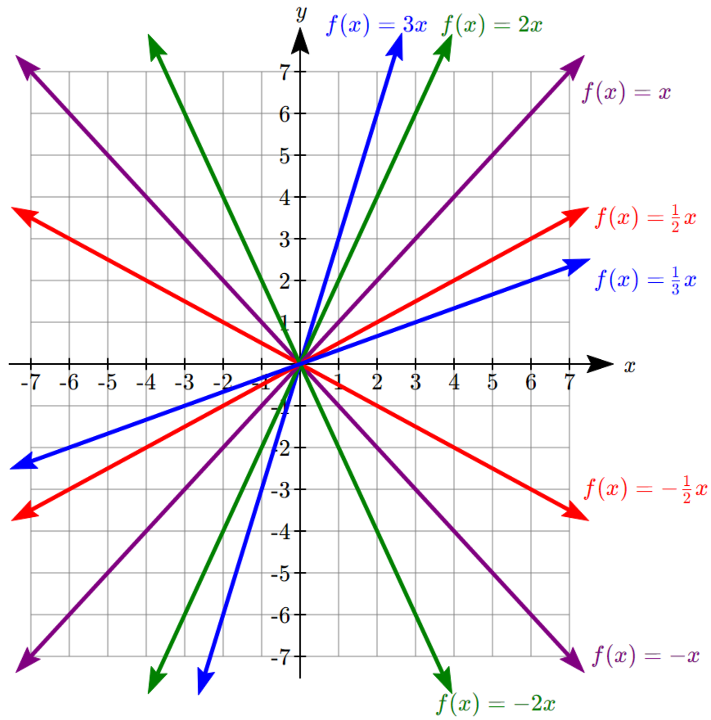


Try it Now

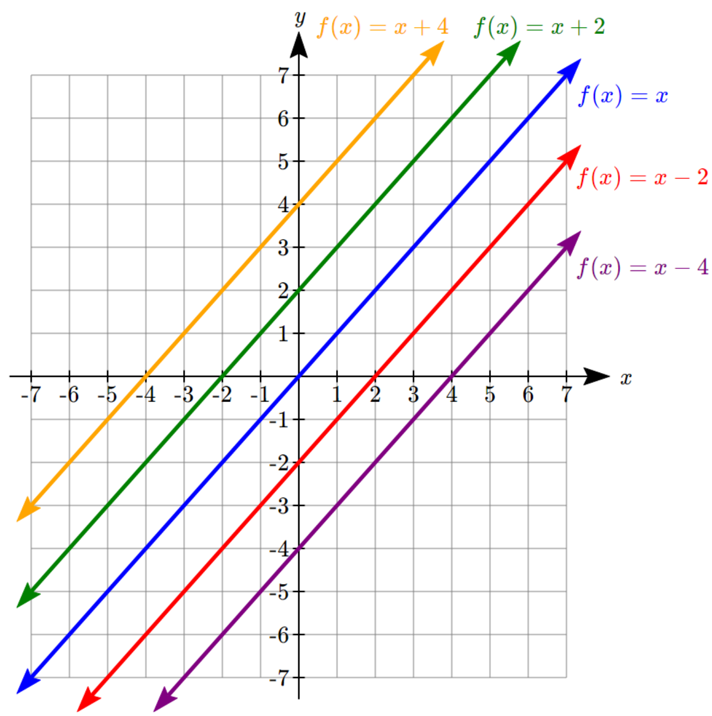
1. Consider that the slope $-\frac{2}{3}$ could also be written as $\frac{2}{-3}$. Using $\frac{2}{-3}$, find another point on the graph that has a negative x value.

Another option for graphing is to use transformations of the identity function $f(x) = x$.

In the equation $f(x) = mx$, the m is acting as the vertical stretch of the identity function. When m is negative, there is also a vertical reflection of the graph. Looking at some examples:



In $f(x) = mx + b$, the b acts as the vertical shift, moving the graph up and down without affecting the slope of the line. Some examples:



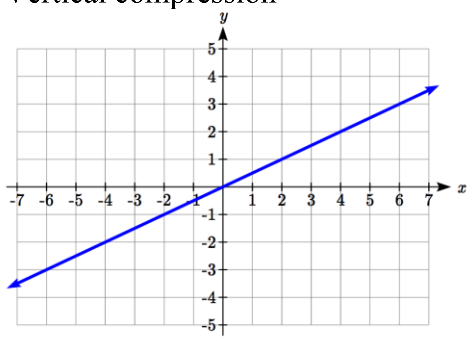
Using Vertical Stretches or Compressions along with Vertical Shifts is another way to look at identifying different types of linear functions. Although this may not be the easiest way for you to graph this type of function, make sure you practice each method.

Example 3

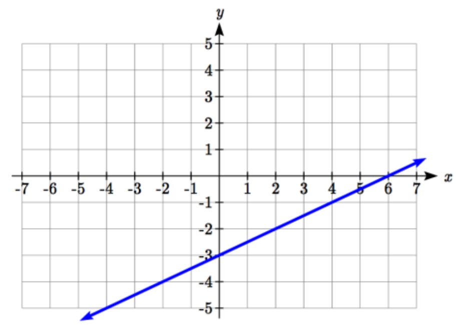
Graph $f(x) = -3 + \frac{1}{2}x$ using transformations.

The equation is the graph of the identity function vertically compressed by ½ and vertically shifted down 3.

Vertical compression



combined with Vertical shift



Notice how this nicely compares to the other method where the vertical intercept is found at (0, -3) and to get to another point we rise (go up vertically) by 1 unit and run (go horizontally) by 2 units to get to the next point (2, -2), and the next one (4, -1). In these three points (0, -3), (2, -2), and (4, -1), the output values change by +1, and the *x* values change by +2, corresponding with the slope m = ½.

Example 4

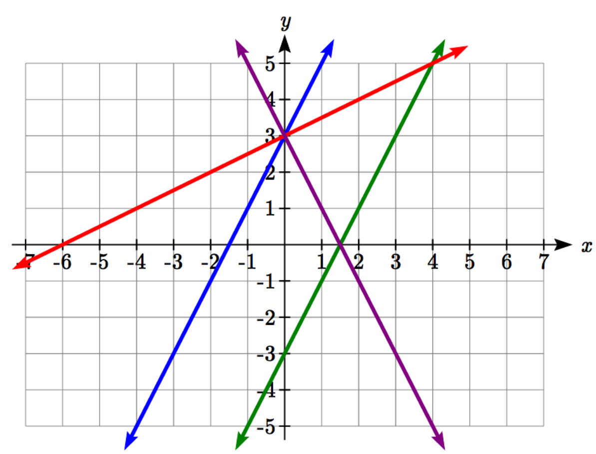
Match each equation with one of the lines in the graph below

$f(x) = 2x + 3$

$g(x) = 2x - 3$

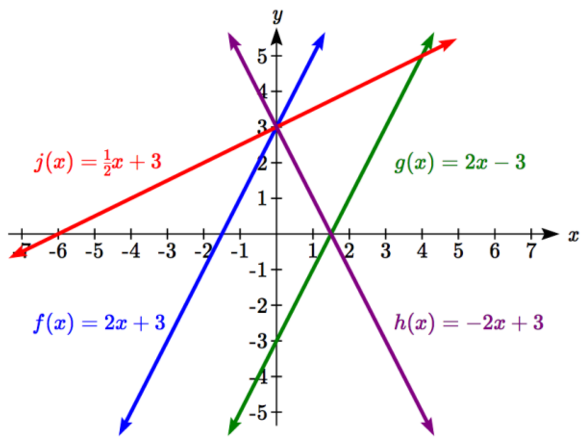
$h(x) = -2x + 3$

$j(x) = \frac{1}{2}x + 3$



Only one graph has a vertical intercept of -3, so we can immediately match that graph with *g(x)*.

For the three graphs with a vertical intercept at 3, only one has a negative slope, so we can match that line with *h(x)*. Of the other two, the steeper line would have a larger slope, so we can match that graph with equation *f(x)*, and the flatter line with the equation *j(x)*.



In addition to understanding the basic behavior of a linear function (increasing or decreasing, recognizing the slope and vertical intercept), it is often helpful to know the horizontal intercept of the function – where it crosses the horizontal axis.

Finding Horizontal Intercepts

The **horizontal intercept** of the function is where the graph crosses the horizontal axis. If a function has a horizontal intercept, you can always find it by solving *f(x)* = 0.

Example 5

Find the horizontal intercept of $f(x) = -3 + \frac{1}{2}x$

Setting the function equal to zero to find what input will put us on the horizontal axis,

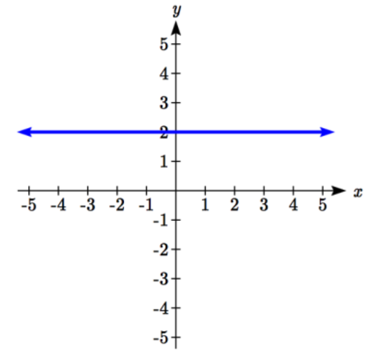
$$0 = -3 + \frac{1}{2}x$$

$$3 = \frac{1}{2}x$$

$$x = 6$$

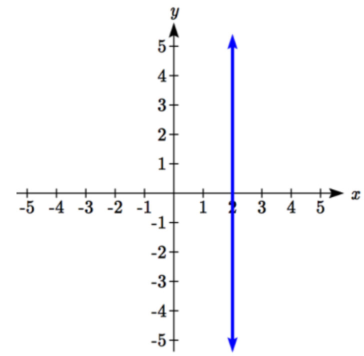
The graph crosses the horizontal axis at (6,0)

There are two special cases of lines: a horizontal line and a vertical line. In a horizontal line like the one graphed to the right, notice that between any two points, the change in the outputs is 0. In the slope equation, the numerator will be 0, resulting in a slope of 0. Using a slope of 0 in the $f(x) = b + mx$, the equation simplifies to $f(x) = b$.



Notice a horizontal line has a vertical intercept, but no horizontal intercept (unless it's the line $f(x) = 0$).

In the case of a vertical line, notice that between any two points, the change in the inputs is zero. In the slope equation, the denominator will be zero, and you may recall that we cannot divide by the zero; the slope of a vertical line is undefined. You might also notice that a vertical line is not a function. To write the equation of vertical line, we simply write input=value, like $x = b$.
Notice a vertical line has a horizontal intercept, but no vertical intercept (unless it's the line $x = 0$).



Horizontal and Vertical Lines

Horizontal lines have equations of the form $f(x) = b$

Vertical lines have equations of the form $x = a$

Example 6

Write an equation for the horizontal line graphed above.

This line would have equation $f(x) = 2$

Example 7

Write an equation for the vertical line graphed above.

This line would have equation $x = 2$

Try it Now

2. Describe the function $f(x) = 6 - 3x$ in terms of transformations of the identity function and find its horizontal intercept.

Parallel and Perpendicular Lines

When two lines are graphed together, the lines will be **parallel** if they are increasing at the same rate – if the rates of change are the same. In this case, the graphs will never cross (unless they're the same line).

Parallel Lines

Two lines are **parallel** if the slopes are equal (or, if both lines are vertical).

In other words, given two linear equations $f(x) = b + m_1 x$ and $g(x) = b + m_2 x$, the lines will be parallel if $m_1 = m_2$.

Example 8

Find a line parallel to $f(x) = 6 + 3x$ that passes through the point (3, 0)

We know the line we're looking for will have the same slope as the given line, $m = 3$. Using this and the given point, we can solve for the new line's vertical intercept:

$g(x) = b + 3x$ then at (3, 0),

$0 = b + 3(3)$

$b = -9$

The line we're looking for is $g(x) = -9 + 3x$

If two lines are not parallel, one other interesting possibility is that the lines are perpendicular, which means the lines form a right angle (90 degree angle – a square corner) where they meet. In this case, the slopes when multiplied together will equal -1. Solving for one slope leads us to the definition:

Perpendicular Lines

Given two linear equations $f(x) = b + m_1 x$ and $g(x) = b + m_2 x$

The lines will be **perpendicular** if $m_1 m_2 = -1$, and so $m_2 = \dfrac{-1}{m_1}$

We often say the slope of a perpendicular line is the "negative reciprocal" of the other line's slope.

Example 9

Find the slope of a line perpendicular to a line with:

a) a slope of 2. b) a slope of -4. c) a slope of $\frac{2}{3}$.

If the original line had slope 2, the perpendicular line's slope would be $m_2 = \frac{-1}{2}$

If the original line had slope -4, the perpendicular line's slope would be $m_2 = \frac{-1}{-4} = \frac{1}{4}$

If the original line had slope $\frac{2}{3}$, the perpendicular line's slope would be $m_2 = \frac{-1}{2/3} = \frac{-3}{2}$

Example 10

Find the equation of a line perpendicular to $f(x) = 6 + 3x$ and passing through the point (3, 0)

The original line has slope $m = 3$. The perpendicular line will have slope $m = \frac{-1}{3}$.

Using this and the given point, we can find the equation for the line.

$g(x) = b - \frac{1}{3}x$ then at (3, 0),

$0 = b - \frac{1}{3}(3)$

$b = 1$

The line we're looking for is $g(x) = 1 - \frac{1}{3}x$.

Try it Now

3. Given the line $h(t) = -4 + 2t$, find an equation for the line passing through (0, 0) that is: a) parallel to *h(t)*. b) perpendicular to *h(t)*.

Example 12

A line passes through the points (-2, 6) and (4, 5). Find the equation of a perpendicular line that passes through the point (4, 5).

From the two given points on the reference line, we can calculate the slope of that line:

$$m_1 = \frac{5-6}{4-(-2)} = \frac{-1}{6}$$

The perpendicular line will have slope

$$m_2 = \frac{-1}{-1/6} = 6$$

We can then solve for the vertical intercept that makes the line pass through the desired point:

$g(x) = b + 6x$ then at (4, 5),

$5 = b + 6(4)$

$b = -19$

Giving the line $g(x) = -19 + 6x$

Intersections of Lines

The graphs of two lines will intersect if they are not parallel. They will intersect at the point that satisfies both equations. To find this point when the equations are given as functions, we can solve for an input value so that $f(x) = g(x)$. In other words, we can set the formulas for the lines equal, and solve for the input that satisfies the equation.

Example 13

Find the intersection of the lines $h(t) = 3t - 4$ and $j(t) = 5 - t$

Setting $h(t) = j(t)$,

$3t - 4 = 5 - t$

$4t = 9$

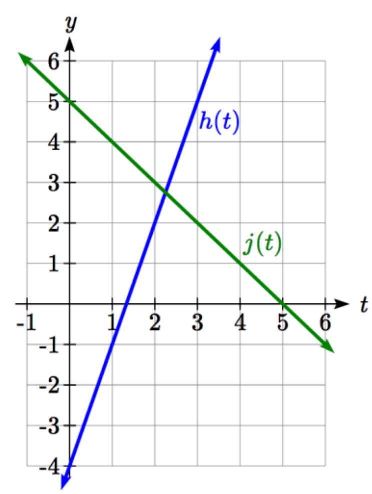
$t = \frac{9}{4}$

This tells us the lines intersect when the input is $\frac{9}{4}$.

We can then find the output value of the intersection point by evaluating either function at this input

$$j\left(\frac{9}{4}\right) = 5 - \frac{9}{4} = \frac{11}{4}$$

These lines intersect at the point $\left(\frac{9}{4}, \frac{11}{4}\right)$. Looking at the graph, this result seems reasonable.



Two parallel lines can also intersect if they happen to be the same line. In that case, they intersect at every point on the lines.

Try it Now

4. Look at the graph in example 13 above and answer the following for the function *h(t)*:
   a. Vertical intercept coordinates
   b. Horizontal intercepts coordinates
   c. Slope
   d. Is *j(t)* parallel or perpendicular to *h(t)* (or neither)
   e. Is *h(t)* an Increasing or Decreasing function (or neither)
   f. Write a transformation description from the identity toolkit function *f(x) = x*

Finding the intersection allows us to answer other questions as well, such as discovering when one function is larger than another.

Example 14

Using the functions from the previous example, for what values of t is $h(t) > j(t)$

To answer this question, it is helpful first to know where the functions are equal, since that is the point where *h(t)* could switch from being greater to smaller than *j(t)* or vice-versa. From the previous example, we know the functions are equal at $t = \frac{9}{4}$.

By examining the graph, we can see that *h(t)*, the function with positive slope, is going to be larger than the other function to the right of the intersection. So $h(t) > j(t)$ when $t > \frac{9}{4}$

Important Topics of this Section
Methods for graphing linear functions
Another name for slope = rise/run
Horizontal intercepts (a,0)
Horizontal lines
Vertical lines
Parallel lines
Perpendicular lines
Intersecting lines

Try it Now Answers

1. (-3,7) found by starting at the vertical intercept, going up 2 units and 3 in the negative horizontal direction. You could have also answered, (-6, 9) or (-9, 11) etc…

2. Vertically stretched by a factor of 3, Vertically flipped (flipped over the x axis), Vertically shifted up by 6 units.
   Horizontal intercept: $6-3x=0$ when $x=2$

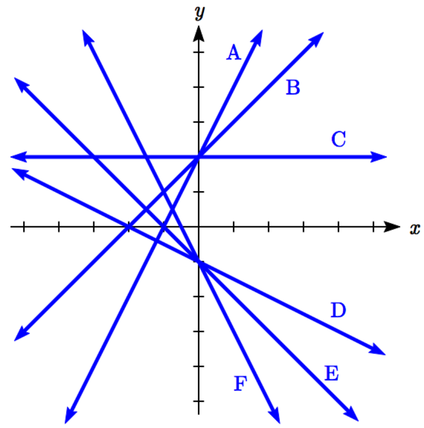
3. Parallel $f(t)=2t$; Perpendicular $g(t)=-\frac{1}{2}t$

4. Given $h(t)=3t-4$
   a. (0,-4)
   b. $\left(\frac{4}{3},0\right)$
   c. Slope 3
   d. Neither parallel nor perpendicular
   e. Increasing function
   f. Given the identity function, vertically stretch by 3 and shift down 4 units.

Section 2.2 Exercises

Match each linear equation with its graph

1. $f(x) = -x - 1$
2. $f(x) = -2x - 1$
3. $f(x) = -\frac{1}{2}x - 1$
4. $f(x) = 2$
5. $f(x) = 2 + x$
6. $f(x) = 3x + 2$



Sketch a line with the given features

7. An *x*-intercept of (-4, 0) and *y*-intercept of (0, -2)
8. An *x*-intercept of (-2, 0) and *y*-intercept of (0, 4)
9. A vertical intercept of (0, 7) and slope $-\frac{3}{2}$
10. A vertical intercept of (0, 3) and slope $\frac{2}{5}$
11. Passing through the points (-6,-2) and (6,-6)
12. Passing through the points (-3,-4) and (3,0)

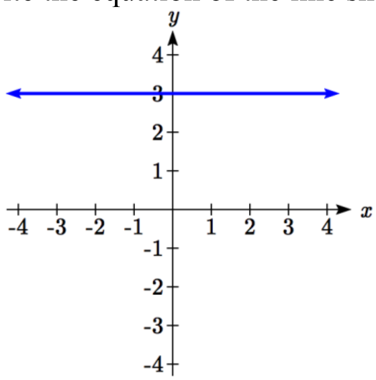
Sketch the graph of each equation

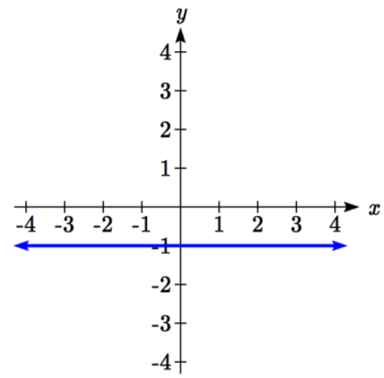
13. $f(x) = -2x - 1$
14. $g(x) = -3x + 2$
15. $h(x) = \frac{1}{3}x + 2$
16. $k(x) = \frac{2}{3}x - 3$
17. $k(t) = 3 + 2t$
18. $p(t) = -2 + 3t$
19. $x = 3$
20. $x = -2$
21. $r(x) = 4$
22. $q(x) = 3$

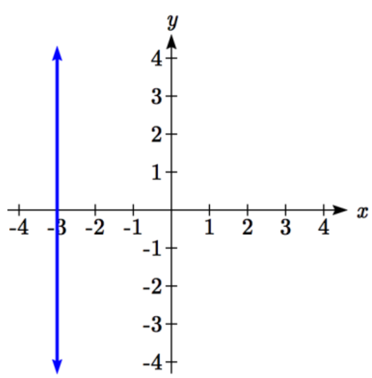
23. If $g(x)$ is the transformation of $f(x)=x$ after a vertical compression by $3/4$, a shift left by 2, and a shift down by 4
    a. Write an equation for $g(x)$
    b. What is the slope of this line?
    c. Find the vertical intercept of this line.

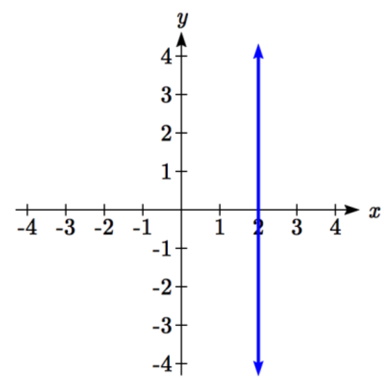
24. If $g(x)$ is the transformation of $f(x)=x$ after a vertical compression by $1/3$, a shift right by 1, and a shift up by 3
    a. Write an equation for $g(x)$
    b. What is the slope of this line?
    c. Find the vertical intercept of this line.

Write the equation of the line shown

25.


26.


27.


28.


Find the horizontal and vertical intercepts of each equation

29. $f(x)=-x+2$

30. $g(x)=2x+4$

31. $h(x)=3x-5$

32. $k(x)=-5x+1$

33. $-2x+5y=20$

34. $7x+2y=56$

Given below are descriptions of two lines. Find the slopes of Line 1 and Line 2. Is each pair of lines parallel, perpendicular or neither?

35. Line 1: Passes through $(0,6)$ and $(3,-24)$
Line 2: Passes through $(-1,19)$ and $(8,-71)$

36. Line 1: Passes through $(-8,-55)$ and $(10,89)$
Line 2: Passes through $(9,-44)$ and $(4,-14)$

37. Line 1: Passes through $(2,3)$ and $(4,-1)$
Line 2: Passes through $(6,3)$ and $(8,5)$

38. Line 1: Passes through $(1,7)$ and $(5,5)$
Line 2: Passes through $(-1,-3)$ and $(1,1)$

39. Line 1: Passes through $(0,5)$ and $(3,3)$
Line 2: Passes through $(1,-5)$ and $(3,-2)$

40. Line 1: Passes through $(2,5)$ and $(5,-1)$
Line 2: Passes through $(-3,7)$ and $(3,-5)$

41. Write an equation for a line parallel to $f(x) = -5x - 3$ and passing through the point (2,-12)

42. Write an equation for a line parallel to $g(x) = 3x - 1$ and passing through the point (4,9)

43. Write an equation for a line perpendicular to $h(t) = -2t + 4$ and passing through the point (-4,-1)

44. Write an equation for a line perpendicular to $p(t) = 3t + 4$ and passing through the point (3,1)

45. Find the point at which the line $f(x) = -2x - 1$ intersects the line $g(x) = -x$

46. Find the point at which the line $f(x) = 2x + 5$ intersects the line $g(x) = -3x - 5$

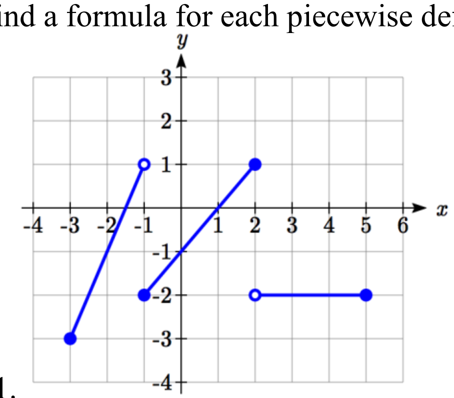
47. Use algebra to find the point at which the line $f(x) = -\frac{4}{5}x + \frac{274}{25}$ intersects the line $h(x) = \frac{9}{4}x + \frac{73}{10}$

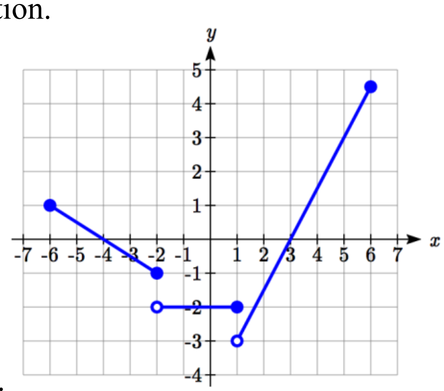
48. Use algebra to find the point at which the line $f(x) = \frac{7}{4}x + \frac{457}{60}$ intersects the line $g(x) = \frac{4}{3}x + \frac{31}{5}$

49. A car rental company offers two plans for renting a car.
Plan A: 30 dollars per day and 18 cents per mile
Plan B: 50 dollars per day with free unlimited mileage
How many miles would you need to drive for plan B to save you money?

50. You're comparing two cell phone companies.
Company A: \$20/month for unlimited talk and text, and \$10/GB for data.
Company B: \$65/month for unlimited talk, text, and data.
Under what circumstances will company A save you money?

Find a formula for each piecewise defined function.

51. 

52. 

53. Sketch an accurate picture of the line having equation $f(x) = 2 - \frac{1}{2}x$. Let c be an unknown constant. [UW]

a. Find the point of intersection between the line you have graphed and the line $g(x) = 1 + cx$; your answer will be a point in the xy plane whose coordinates involve the unknown c.
b. Find c so that the intersection point in (a) has x-coordinate 10.
c. Find c so that the intersection point in (a) lies on the x-axis.

Section 2.3 Modeling with Linear Functions

When modeling scenarios with a linear function and solving problems involving quantities changing linearly, we typically follow the same problem solving strategies that we would use for any type of function:

Problem solving strategy

1) Identify changing quantities, and then carefully and clearly define descriptive variables to represent those quantities. When appropriate, sketch a picture or define a coordinate system.
2) Carefully read the problem to identify important information. Look for information giving values for the variables, or values for parts of the functional model, like slope and initial value.
3) Carefully read the problem to identify what we are trying to find, identify, solve, or interpret.
4) Identify a solution pathway from the provided information to what we are trying to find. Often this will involve checking and tracking units, building a table or even finding a formula for the function being used to model the problem.
5) When needed, find a formula for the function.
6) Solve or evaluate using the formula you found for the desired quantities.
7) Reflect on whether your answer is reasonable for the given situation and whether it makes sense mathematically.
8) Clearly convey your result using appropriate units, and answer in full sentences when appropriate.

Example 1

Emily saved up $3500 for her summer visit to Seattle. She anticipates spending $400 each week on rent, food, and fun. Find and interpret the horizontal intercept and determine a reasonable domain and range for this function.

In the problem, there are two changing quantities: time and money. The amount of money she has remaining while on vacation depends on how long she stays. We can define our variables, including units.
Output: M, money remaining, in dollars
Input: t, time, in weeks

Reading the problem, we identify two important values. The first, $3500, is the initial value for M. The other value appears to be a rate of change – the units of dollars per week match the units of our output variable divided by our input variable. She is spending money each week, so you should recognize that the amount of money remaining is decreasing each week and the slope is negative.

To answer the first question, looking for the horizontal intercept, it would be helpful to have an equation modeling this scenario. Using the intercept and slope provided in the problem, we can write the equation: $M(t) = 3500 - 400t$.

To find the horizontal intercept, we set the output to zero, and solve for the input:

$0 = 3500 - 400t$

$t = \dfrac{3500}{400} = 8.75$

The horizontal intercept is 8.75 weeks. Since this represents the input value where the output will be zero, interpreting this, we could say: Emily will have no money left after 8.75 weeks.

When modeling any real life scenario with functions, there is typically a limited domain over which that model will be valid – almost no trend continues indefinitely. In this case, it certainly doesn't make sense to talk about input values less than zero. It is also likely that this model is not valid after the horizontal intercept (unless Emily's going to start using a credit card and go into debt).

The domain represents the set of input values and so the reasonable domain for this function is $0 \le t \le 8.75$.

However, in a real world scenario, the rental might be weekly or nightly. She may not be able to stay a partial week and so all options should be considered. Emily could stay in Seattle for 0 to 8 full weeks (and a couple of days), but would have to go into debt to stay 9 full weeks, so restricted to whole weeks, a reasonable domain without going in to debt would be $0 \le t \le 8$, or $0 \le t \le 9$ if she went into debt to finish out the last week.

The range represents the set of output values and she starts with \$3500 and ends with \$0 after 8.75 weeks so the corresponding range is $0 \le M(t) \le 3500$. If we limit the rental to whole weeks, however, the range would change. If she left after 8 weeks because she didn't have enough to stay for a full 9 weeks, she would have $M(8) = 3500 - 400(8) =$ \$300 dollars left after 8 weeks, giving a range of $300 \le M(t) \le 3500$. If she wanted to stay the full 9 weeks she would be \$100 in debt giving a range of $-100 \le M(t) \le 3500$.

Most importantly remember that domain and range are tied together, and what ever you decide is most appropriate for the domain (the independent variable) will dictate the requirements for the range (the dependent variable).

Try it Now

1. A database manager is loading a large table from backups. Getting impatient, she notices 1.2 million rows had been loaded. Ten minutes later, 2.5 million rows had been loaded. How much longer will she have to wait for all 80 million rows to load?

Example 2

Jamal is choosing between two moving companies. The first, U-Haul, charges an up-front fee of \$20, then 59 cents a mile. The second, Budget, charges an up-front fee of \$16, then 63 cents a mile[3]. When will U-Haul be the better choice for Jamal?

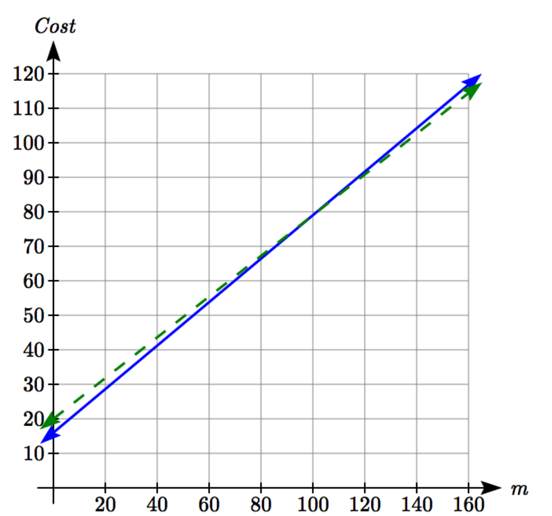
The two important quantities in this problem are the cost, and the number of miles that are driven. Since we have two companies to consider, we will define two functions:

Input: m, miles driven
Outputs:
$Y(m)$: cost, in dollars, for renting from U-Haul
$B(m)$: cost, in dollars, for renting from Budget

Reading the problem carefully, it appears that we were given an initial cost and a rate of change for each company. Since our outputs are measured in dollars but the costs per mile given in the problem are in cents, we will need to convert these quantities to match our desired units: \$0.59 a mile for U-Haul, and \$0.63 a mile for Budget.

Looking to what we're trying to find, we want to know when U-Haul will be the better choice. Since all we have to make that decision from is the costs, we are looking for when U-Haul will cost less, or when $Y(m) < B(m)$. The solution pathway will lead us to find the equations for the two functions, find the intersection, then look to see where the $Y(m)$ function is smaller. Using the rates of change and initial charges, we can write the equations:

$$Y(m) = 20 + 0.59m$$
$$B(m) = 16 + 0.63m$$



These graphs are sketched to the right, with $Y(m)$ drawn dashed.

To find the intersection, we set the equations equal and solve:

$$Y(m) = B(m)$$
$$20 + 0.59m = 16 + 0.63m$$
$$4 = 0.04m$$
$$m = 100$$

This tells us that the cost from the two companies will be the same if 100 miles are driven. Either by looking at the graph, or noting that $Y(m)$ is growing at a slower rate, we can conclude that U-Haul will be the cheaper price when more than 100 miles are driven.

[3] Rates retrieved Aug 2, 2010 from http://www.budgettruck.com and http://www.uhaul.com/

Example 3

A town's population has been growing linearly. In 2004 the population was 6,200. By 2009 the population had grown to 8,100. If this trend continues,
a. Predict the population in 2013
b. When will the population reach 15000?

The two changing quantities are the population and time. While we could use the actual year value as the input quantity, doing so tends to lead to very ugly equations, since the vertical intercept would correspond to the year 0, more than 2000 years ago!
To make things a little nicer, and to make our lives easier too, we will define our input as years since 2004:
Input: t, years since 2004
Output: $P(t)$, the town's population

The problem gives us two input-output pairs. Converting them to match our defined variables, the year 2004 would correspond to $t = 0$, giving the point (0, 6200). Notice that through our clever choice of variable definition, we have "given" ourselves the vertical intercept of the function. The year 2009 would correspond to $t = 5$, giving the point (5, 8100).

To predict the population in 2013 ($t = 9$), we would need an equation for the population. Likewise, to find when the population would reach 15000, we would need to solve for the input that would provide an output of 15000. Either way, we need an equation. To find it, we start by calculating the rate of change:

$$m = \frac{8100 - 6200}{5 - 0} = \frac{1900}{5} = 380 \text{ people per year}$$

Since we already know the vertical intercept of the line, we can immediately write the equation:

$$P(t) = 6200 + 380t$$

To predict the population in 2013, we evaluate our function at $t = 9$

$$P(9) = 6200 + 380(9) = 9620$$

If the trend continues, our model predicts a population of 9,620 in 2013.

To find when the population will reach 15,000, we can set $P(t) = 15000$ and solve for t.

$$15000 = 6200 + 380t$$

$$8800 = 380t$$

$$t \approx 23.158$$

Our model predicts the population will reach 15,000 in a little more than 23 years after 2004, or somewhere around the year 2027.

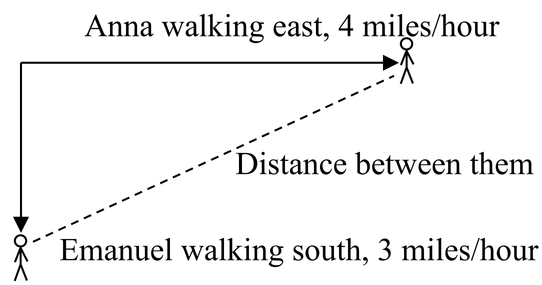
Example 4

Anna and Emanuel start at the same intersection. Anna walks east at 4 miles per hour while Emanuel walks south at 3 miles per hour. They are communicating with a two-way radio with a range of 2 miles. How long after they start walking will they fall out of radio contact?

In essence, we can partially answer this question by saying they will fall out of radio contact when they are 2 miles apart, which leads us to ask a new question: how long will it take them to be 2 miles apart?

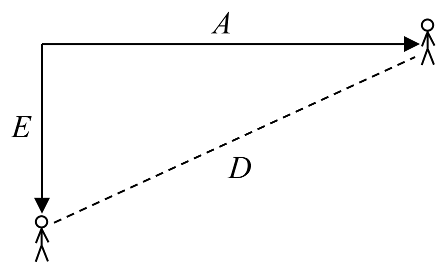
In this problem, our changing quantities are time and the two peoples' positions, but ultimately we need to know how long will it take for them to be 2 miles apart. We can see that time will be our input variable, so we'll define
Input: t, time in hours.

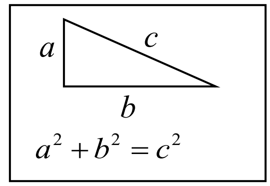
Since it is not obvious how to define our output variables, we'll start by drawing a picture.



Because of the complexity of this question, it may be helpful to introduce some intermediary variables. These are quantities that we aren't directly interested in, but seem important to the problem. For this problem, Anna's and Emanuel's distances from the starting point seem important. To notate these, we are going to define a coordinate system, putting the "starting point" at the intersection where they both started, then we're going to introduce a variable, A, to represent Anna's position, and define it to be a measurement from the starting point, in the eastward direction. Likewise, we'll introduce a variable, E, to represent Emanuel's position, measured from the starting point in the southward direction. Note that in defining the coordinate system we specified both the origin, or starting point, of the measurement, as well as the direction of measure.

While we're at it, we'll define a third variable, D, to be the measurement of the distance between Anna and Emanuel. Showing the variables on the picture is often helpful: Looking at the variables on the picture, we remember we need to know how long it takes for D, the distance between them, to equal 2 miles.





Seeing this picture we remember that in order to find the distance between the two, we can use the Pythagorean Theorem, a property of right triangles.

From here, we can now look back at the problem for relevant information. Anna is walking 4 miles per hour, and Emanuel is walking 3 miles per hour, which are rates of change. Using those, we can write formulas for the distance each has walked.

They both start at the same intersection and so when $t = 0$, the distance travelled by each person should also be 0, so given the rate for each, and the initial value for each, we get:

$A(t) = 4t$

$E(t) = 3t$

Using the Pythagorean theorem we get:

$D(t)^2 = A(t)^2 + E(t)^2$ substitute in the function formulas

$D(t)^2 = (4t)^2 + (3t)^2 = 16t^2 + 9t^2 = 25t^2$ solve for *D(t)* using the square root

$D(t) = \pm\sqrt{25t^2} = \pm 5|t|$

Since in this scenario we are only considering positive values of t and our distance *D(t)* will always be positive, we can simplify this answer to $D(t) = 5t$

Interestingly, the distance between them is also a linear function. Using it, we can now answer the question of when the distance between them will reach 2 miles:

$D(t) = 2$

$5t = 2$

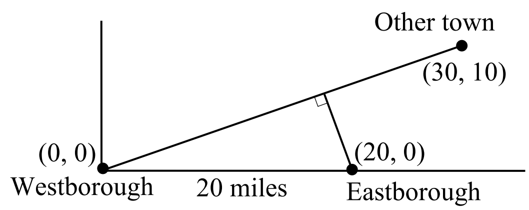
$t = \frac{2}{5} = 0.4$

They will fall out of radio contact in 0.4 hours, or 24 minutes.

Example 5

There is currently a straight road leading from the town of Westborough to a town 30 miles east and 10 miles north. Partway down this road, it junctions with a second road, perpendicular to the first, leading to the town of Eastborough. If the town of Eastborough is located 20 miles directly east of the town of Westborough, how far is the road junction from Westborough?

It might help here to draw a picture of the situation. It would then be helpful to introduce a coordinate system. While we could place the origin anywhere, placing it at Westborough seems convenient. This puts the other town at coordinates (30, 10), and Eastborough at (20, 0).



Using this point along with the origin, we can find the slope of the line from Westborough to the other town: $m = \dfrac{10-0}{30-0} = \dfrac{1}{3}$. This gives the equation of the road from Westborough to the other town to be $W(x) = \dfrac{1}{3}x$.

From this, we can determine the perpendicular road to Eastborough will have slope $m = -3$. Since the town of Eastborough is at the point (20, 0), we can find the equation:

$E(x) = -3x + b$ plug in the point (20, 0)

$0 = -3(20) + b$

$b = 60$

$E(x) = -3x + 60$

We can now find the coordinates of the junction of the roads by finding the intersection of these lines. Setting them equal,

$\dfrac{1}{3}x = -3x + 60$

$\dfrac{10}{3}x = 60$

$10x = 180$

$x = 18$ Substituting this back into *W(x)*

$y = W(18) = \dfrac{1}{3}(18) = 6$

The roads intersect at the point (18, 6). Using the distance formula, we can now find the distance from Westborough to the junction:

$dist = \sqrt{(18-0)^2 + (6-0)^2} \approx 18.934$ miles.

Important Topics of this Section
The problem solving process 1) Identify changing quantities, and then carefully and clearly define descriptive variables to represent those quantities. When appropriate, sketch a picture or define a coordinate system. 2) Carefully read the problem to identify important information. Look for information giving values for the variables, or values for parts of the functional model, like slope and initial value. 3) Carefully read the problem to identify what we are trying to find, identify, solve, or interpret. 4) Identify a solution pathway from the provided information to what we are trying to find. Often this will involve checking and tracking units, building a table or even finding a formula for the function being used to model the problem. 5) When needed, find a formula for the function. 6) Solve or evaluate using the formula you found for the desired quantities. 7) Reflect on whether your answer is reasonable for the given situation and whether it makes sense mathematically. 8) Clearly convey your result using appropriate units, and answer in full sentences when appropriate.

Try it Now

1. Letting t be the number of minutes since she got impatient, and N be the number rows loaded, in millions, we have two points: (0, 1.2) and (10, 2.5).

   The slope is $m = \dfrac{2.5-1.2}{10-0} = \dfrac{1.3}{10} = 0.13$ million rows per minute.
   We know the N intercept, so we can write the equation:
   $N = 0.13t + 1.2$

   To determine how long she will have to wait, we need to solve for when $N = 80$.
   $N = 0.13t + 1.2 = 80$
   $0.13t = 78.8$
   $t = \dfrac{78.8}{0.13} \approx 606$. She'll have to wait another 606 minutes, about 10 hours.

Section 2.3 Exercises

1. In 2004, a school population was 1001. By 2008 the population had grown to 1697. Assume the population is changing linearly.
    a. How much did the population grow between the year 2004 and 2008?
    b. How long did it take the population to grow from 1001 students to 1697 students?
    c. What is the average population growth per year?
    d. What was the population in the year 2000?
    e. Find an equation for the population, P, of the school t years after 2000.
    f. Using your equation, predict the population of the school in 2011.

2. In 2003, a town's population was 1431. By 2007 the population had grown to 2134. Assume the population is changing linearly.
    a. How much did the population grow between the year 2003 and 2007?
    b. How long did it take the population to grow from 1431 people to 2134?
    c. What is the average population growth per year?
    d. What was the population in the year 2000?
    e. Find an equation for the population, P, of the town t years after 2000.
    f. Using your equation, predict the population of the town in 2014.

3. A phone company has a monthly cellular plan where a customer pays a flat monthly fee and then a certain amount of money per minute used on the phone. If a customer uses 410 minutes, the monthly cost will be \$71.50. If the customer uses 720 minutes, the monthly cost will be \$118.
    a. Find a linear equation for the monthly cost of the cell plan as a function of x, the number of monthly minutes used.
    b. Interpret the slope and vertical intercept of the equation.
    c. Use your equation to find the total monthly cost if 687 minutes are used.

4. A phone company has a monthly cellular data plan where a customer pays a flat monthly fee and then a certain amount of money per megabyte (MB) of data used on the phone. If a customer uses 20 MB, the monthly cost will be \$11.20. If the customer uses 130 MB, the monthly cost will be \$17.80.
    a. Find a linear equation for the monthly cost of the data plan as a function of x, the number of MB used.
    b. Interpret the slope and vertical intercept of the equation.
    c. Use your equation to find the total monthly cost if 250 MB are used.

5. In 1991, the moose population in a park was measured to be 4360. By 1999, the population was measured again to be 5880. If the population continues to change linearly,
   a. Find a formula for the moose population, P.
   b. What does your model predict the moose population to be in 2003?

6. In 2003, the owl population in a park was measured to be 340. By 2007, the population was measured again to be 285. If the population continues to change linearly,
   a. Find a formula for the owl population, P.
   b. What does your model predict the owl population to be in 2012?

7. The Federal Helium Reserve held about 16 billion cubic feet of helium in 2010, and is being depleted by about 2.1 billion cubic feet each year.
   a. Give a linear equation for the remaining federal helium reserves, R, in terms of t, the number of years since 2010.
   b. In 2015, what will the helium reserves be?
   c. If the rate of depletion doesn't change, when will the Federal Helium Reserve be depleted?

8. Suppose the world's current oil reserves are 1820 billion barrels. If, on average, the total reserves is decreasing by 25 billion barrels of oil each year:
   a. Give a linear equation for the remaining oil reserves, R, in terms of t, the number of years since now.
   b. Seven years from now, what will the oil reserves be?
   c. If the rate of depletion isn't change, when will the world's oil reserves be depleted?

9. You are choosing between two different prepaid cell phone plans. The first plan charges a rate of 26 cents per minute. The second plan charges a monthly fee of \$19.95 *plus* 11 cents per minute. How many minutes would you have to use in a month in order for the second plan to be preferable?

10. You are choosing between two different window washing companies. The first charges \$5 per window. The second charges a base fee of \$40 plus \$3 per window. How many windows would you need to have for the second company to be preferable?

11. When hired at a new job selling jewelry, you are given two pay options:
    Option A: Base salary of \$17,000 a year, with a commission of 12% of your sales
    Option B: Base salary of \$20,000 a year, with a commission of 5% of your sales
    How much jewelry would you need to sell for option A to produce a larger income?

12. When hired at a new job selling electronics, you are given two pay options:
Option A: Base salary of \$14,000 a year, with a commission of 10% of your sales
Option B: Base salary of \$19,000 a year, with a commission of 4% of your sales
How much electronics would you need to sell for option A to produce a larger income?

13. Find the area of a triangle bounded by the *y* axis, the line $f(x)=9-\frac{6}{7}x$, and the line perpendicular to $f(x)$ that passes through the origin.

14. Find the area of a triangle bounded by the *x* axis, the line $f(x)=12-\frac{1}{3}x$, and the line perpendicular to $f(x)$ that passes through the origin.

15. Find the area of a parallelogram bounded by the *y* axis, the line $x=3$, the line $f(x)=1+2x$, and the line parallel to $f(x)$ passing through (2, 7)

16. Find the area of a parallelogram bounded by the *x* axis, the line $g(x)=2$, the line $f(x)=3x$, and the line parallel to $f(x)$ passing through (6, 1)

17. If $b>0$ and $m<0$, then the line $f(x)=b+mx$ cuts off a triangle from the first quadrant. Express the area of that triangle in terms of *m* and *b*. [UW]

18. Find the value of *m* so the lines $f(x)=mx+5$ and $g(x)=x$ and the *y*-axis form a triangle with an area of 10. [UW]

19. The median home values in Mississippi and Hawaii (adjusted for inflation) are shown below. If we assume that the house values are changing linearly,

Year	**Mississippi**	**Hawaii**
1950	25200	74400
2000	71400	272700

a. In which state have home values increased at a higher rate?
b. If these trends were to continue, what would be the median home value in Mississippi in 2010?
c. If we assume the linear trend existed before 1950 and continues after 2000, the two states' median house values will be (or were) equal in what year? (The answer might be absurd)

20. The median home value ins Indiana and Alabama (adjusted for inflation) are shown below. If we assume that the house values are changing linearly,

Year	Indiana	Alabama
1950	37700	27100
2000	94300	85100

    a. In which state have home values increased at a higher rate?
    b. If these trends were to continue, what would be the median home value in Indiana in 2010?
    c. If we assume the linear trend existed before 1950 and continues after 2000, the two states' median house values will be (or were) equal in what year? (The answer might be absurd)

21. Pam is taking a train from the town of Rome to the town of Florence. Rome is located 30 miles due West of the town of Paris. Florence is 25 miles East, and 45 miles North of Rome. On her trip, how close does Pam get to Paris? [UW]

22. You're flying from Joint Base Lewis-McChord (JBLM) to an undisclosed location 226 km south and 230 km east. Mt. Rainier is located approximately 56 km east and 40 km south of JBLM. If you are flying at a constant speed of 800 km/hr, how long after you depart JBLM will you be the closest to Mt. Rainier?

Section 2.4 Fitting Linear Models to Data

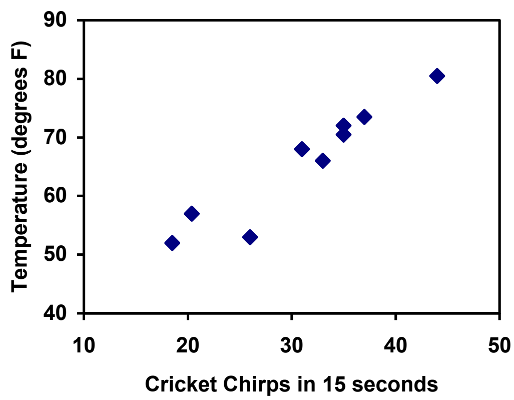
In the real world, rarely do things follow trends perfectly. When we expect the trend to behave linearly, or when inspection suggests the trend is behaving linearly, it is often desirable to find an equation to approximate the data. Finding an equation to approximate the data helps us understand the behavior of the data and allows us to use the linear model to make predictions about the data, inside and outside of the data range.

Example 1

The table below shows the number of cricket chirps in 15 seconds, and the air temperature, in degrees Fahrenheit[4]. Plot this data, and determine whether the data appears to be linearly related.

chirps	44	35	20.4	33	31	35	18.5	37	26
Temp	80.5	70.5	57	66	68	72	52	73.5	53

Plotting this data, it appears there may be a trend, and that the trend appears roughly linear, though certainly not perfectly so.



The simplest way to find an equation to approximate this data is to try to "eyeball" a line that seems to fit the data pretty well, then find an equation for that line based on the slope and intercept.

You can see from the trend in the data that the number of chirps increases as the temperature increases. As you consider a function for this data you should know that you are looking at an increasing function or a function with a positive slope.

[4] Selected data from http://classic.globe.gov/fsl/scientistsblog/2007/10/. Retrieved Aug 3, 2010

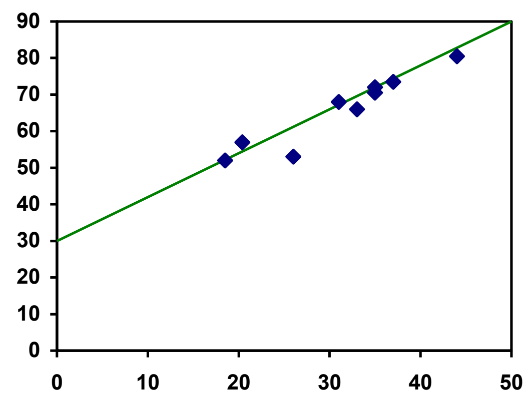
Flashback

1. a. What descriptive variables would you choose to represent Temperature & Chirps?
   b. Which variable is the independent variable and which is the dependent variable?
   c. Based on this data and the graph, what is a reasonable domain & range?
   d. Based on the data alone, is this function one-to-one, explain?

Example 2

Using the table of values from the previous example, find a linear function that fits the data by "eyeballing" a line that seems to fit.

On a graph, we could try sketching in a line. Note the scale on the axes have been adjusted to start at zero to include the vertical axis and vertical intercept in the graph.



Using the starting and ending points of our "hand drawn" line, points (0, 30) and (50, 90), this graph has a slope of $m = \frac{60}{50} = 1.2$ and a vertical intercept at 30, giving an equation of

$$T(c) = 30 + 1.2c$$

where c is the number of chirps in 15 seconds, and $T(c)$ is the temperature in degrees Fahrenheit.

This linear equation can then be used to approximate the solution to various questions we might ask about the trend. While the data does not perfectly fall on the linear equation, the equation is our best guess as to how the relationship will behave outside of the values we have data for. There is a difference, though, between making predictions inside the domain and range of values we have data for, and outside that domain and range.

Interpolation and Extrapolation
Interpolation: When we predict a value inside the domain and range of the data **Extrapolation:** When we predict a value outside the domain and range of the data

For the Temperature as a function of chirps in our hand drawn model above,

- Interpolation would occur if we used our model to predict temperature when the values for chirps are between 18.5 and 44.
- Extrapolation would occur if we used our model to predict temperature when the values for chirps are less than 18.5 or greater than 44.

Example 3

a) Would predicting the temperature when crickets are chirping 30 times in 15 seconds be interpolation or extrapolation? Make the prediction, and discuss if it is reasonable.

b) Would predicting the number of chirps crickets will make at 40 degrees be interpolation or extrapolation? Make the prediction, and discuss if it is reasonable.

With our cricket data, our number of chirps in the data provided varied from 18.5 to 44. A prediction at 30 chirps per 15 seconds is inside the domain of our data, so would be interpolation. Using our model:
$T(30) = 30 + 1.2(30) = 66$ degrees.

Based on the data we have, this value seems reasonable.

The temperature values varied from 52 to 80.5. Predicting the number of chirps at 40 degrees is extrapolation since 40 is outside the range of our data. Using our model:

$$40 = 30 + 1.2c$$

$$10 = 1.2c$$

$$c \approx 8.33$$

Our model predicts the crickets would chirp 8.33 times in 15 seconds. While this might be possible, we have no reason to believe our model is valid outside the domain and range. In fact, generally crickets stop chirping altogether below around 50 degrees.

When our model no longer applies after some point, it is sometimes called **model breakdown**.

Try it Now
1. What temperature would you predict if you counted 20 chirps in 15 seconds?

Fitting Lines with Technology

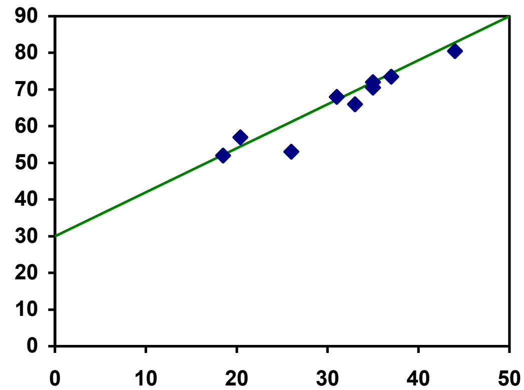
While eyeballing a line works reasonably well, there are statistical techniques for fitting a line to data that minimize the differences between the line and data values[5]. This technique is called **least-square regression**, and can be computed by many graphing calculators, spreadsheet software like Excel or Google Docs, statistical software, and many web-based calculators.

[5] Technically, the method minimizes the sum of the squared differences in the vertical direction between the line and the data values.

Example 4

Find the least-squares regression line using the cricket chirp data from above.

Using the cricket chirp data from earlier, with technology we obtain the equation:
$T(c) = 30.281 + 1.143c$



Notice that this line is quite similar to the equation we "eyeballed", but should fit the data better. Notice also that using this equation would change our prediction for the temperature when hearing 30 chirps in 15 seconds from 66 degrees to:
$T(30) = 30.281 + 1.143(30) = 64.571 \approx 64.6$ degrees.

Most calculators and computer software will also provide you with the **correlation coefficient**, a measure of how closely the line fits the data.

Correlation Coefficient

The **correlation coefficient** is a value, r, between -1 and 1.

$r > 0$ suggests a positive (increasing) relationship

$r < 0$ suggests a negative (decreasing) relationship

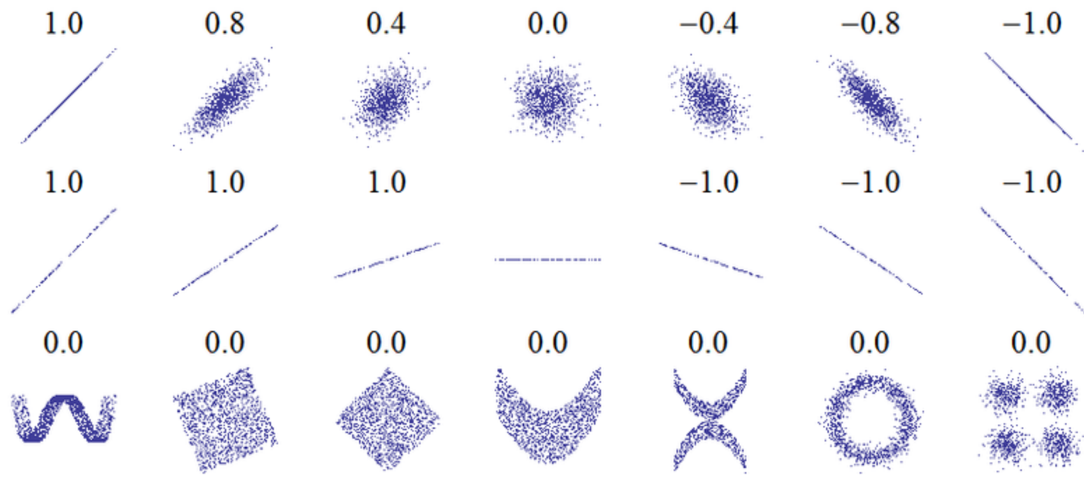
The closer the value is to 0, the more scattered the data

The closer the value is to 1 or -1, the less scattered the data is

The correlation coefficient provides an easy way to get some idea of how close to a line the data falls.

We should only compute the correlation coefficient for data that follows a linear pattern; if the data exhibits a non-linear pattern, the correlation coefficient is meaningless. To get a sense for the relationship between the value of r and the graph of the data, here are some large data sets with their correlation coefficients:

Examples of Correlation Coefficient Values



[6]

Example 5

Calculate the correlation coefficient for our cricket data.

Because the data appears to follow a linear pattern, we can use technology to calculate $r = 0.9509$. Since this value is very close to 1, it suggests a strong increasing linear relationship.

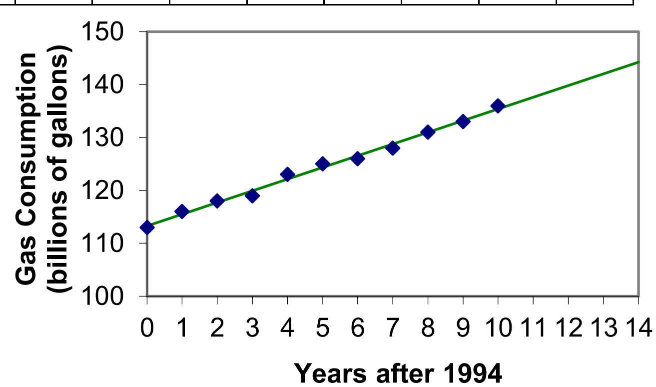
Example 6

Gasoline consumption in the US has been increasing steadily. Consumption data from 1994 to 2004 is shown below.[7] Determine if the trend is linear, and if so, find a model for the data. Use the model to predict the consumption in 2008.

Year	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04
Consumption (billions of gallons)	113	116	118	119	123	125	126	128	131	133	136

To make things simpler, a new input variable is introduced, *t*, representing years since 1994.

Using technology, the correlation coefficient was calculated to be 0.9965, suggesting a very strong increasing linear trend.



[6] http://en.wikipedia.org/wiki/File:Correlation_examples.png
[7] http://www.bts.gov/publications/national_transportation_statistics/2005/html/table_04_10.html

The least-squares regression equation is:
$C(t) = 113.318 + 2.209t$.

Using this to predict consumption in 2008 ($t = 14$),
$C(14) = 113.318 + 2.209(14) = 144.244$ billions of gallons

The model predicts 144.244 billion gallons of gasoline will be consumed in 2008.

Try it Now

2. Use the model created by technology in example 6 to predict the gas consumption in 2011. Is this an interpolation or an extrapolation?

Important Topics of this Section
Fitting linear models to data by hand
Fitting linear models to data using technology
Interpolation
Extrapolation
Correlation coefficient

Flashback Answers

1. a. T = Temperature, C = Chirps (answers may vary)
   b. Independent (Chirps) , Dependent (Temperature)
   c. Reasonable Domain (18.5, 44) , Reasonable Range (52, 80.5) (answers may vary)
   d. NO, it is not one-to-one, there are two different output values for 35 chirps.

Try it Now Answers

1. 54 degrees Fahrenheit
2. 150.871 billion gallons; extrapolation

Section 2.4 Exercises

1. The following is data for the first and second quiz scores for 8 students in a class. Plot the points, then sketch a line that fits the data.

First Quiz	11	20	24	25	33	42	46	49
Second Quiz	10	16	23	28	30	39	40	49

2. Eight students were asked to estimate their score on a 10 point quiz. Their estimated and actual scores are given. Plot the points, then sketch a line that fits the data.

Predicted	5	7	6	8	10	9	10	7
Actual	6	6	7	8	9	9	10	6

Based on each set of data given, calculate the regression line using your calculator or other technology tool, and determine the correlation coefficient.

3.

x	y
5	4
7	12
10	17
12	22
15	24

4.

x	y
8	23
15	41
26	53
31	72
56	103

5.

x	y
3	21.9
4	22.22
5	22.74
6	22.26
7	20.78
8	17.6
9	16.52
10	18.54
11	15.76
12	13.68
13	14.1
14	14.02
15	11.94
16	12.76
17	11.28
18	9.1

6.

x	y
4	44.8
5	43.1
6	38.8
7	39
8	38
9	32.7
10	30.1
11	29.3
12	27
13	25.8
14	24.7
15	22
16	20.1
17	19.8
18	16.8

7. A regression was run to determine if there is a relationship between hours of TV watched per day (x) and number of situps a person can do (y). The results of the regression are given below. Use this to predict the number of situps a person who watches 11 hours of TV can do.

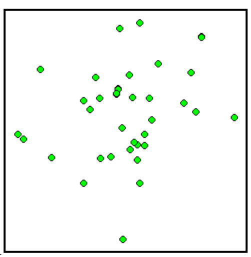
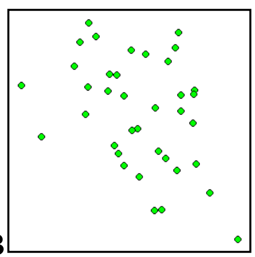
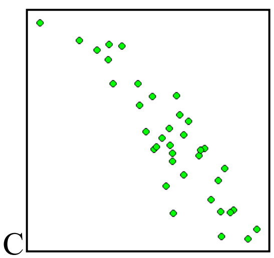
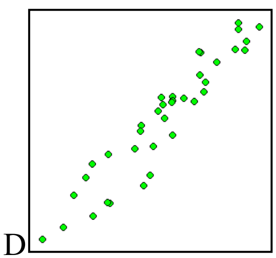
    $y=ax+b$
    $a=-1.341$
    $b=32.234$
    $r^2=0.803$
    $r=-0.896$

8. A regression was run to determine if there is a relationship between the diameter of a tree (x, in inches) and the tree's age (y, in years). The results of the regression are given below. Use this to predict the age of a tree with diameter 10 inches.

    $y=ax+b$
    $a=6.301$
    $b=-1.044$
    $r^2=0.940$
    $r=0.970$

Match each scatterplot shown below with one of the four specified correlations.

9. $r = 0.95$ 10. $r = -0.89$ 11. $r = 0.26$ 12. $r = -0.39$

A 
B 
C 
D 

13. The US census tracks the percentage of persons 25 years or older who are college graduates. That data for several years is given below. Determine if the trend appears linear. If so and the trend continues, in what year will the percentage exceed 35%?

Year	1990	1992	1994	1996	1998	2000	2002	2004	2006	2008
Percent Graduates	21.3	21.4	22.2	23.6	24.4	25.6	26.7	27.7	28	29.4

14. The US import of wine (in hectoliters) for several years is given below. Determine if the trend appears linear. If so and the trend continues, in what year will imports exceed 12,000 hectoliters?

Year	1992	1994	1996	1998	2000	2002	2004	2006	2008	2009
Imports	2665	2688	3565	4129	4584	5655	6549	7950	8487	9462

Section 2.5 Absolute Value Functions

So far in this chapter we have been studying the behavior of linear functions. The Absolute Value Function is a piecewise-defined function made up of two linear functions. The name, Absolute Value Function, should be familiar to you from Section 1.2. In its basic form $f(x)=|x|$ it is one of our toolkit functions.

Absolute Value Function

The absolute value function can be defined as

$$f(x)=|x|=\begin{cases} x & if \quad x \geq 0 \\ -x & if \quad x < 0 \end{cases}$$

The absolute value function is commonly used to determine the distance between two numbers on the number line. Given two values a and b, then $|a-b|$ will give the distance, a positive quantity, between these values, regardless of which value is larger.

Example 1

Describe all values, x, within a distance of 4 from the number 5.

We want the distance between x and 5 to be less than or equal to 4. The distance can be represented using the absolute value, giving the expression

$$|x-5| \leq 4$$

Example 2

A 2010 poll reported 78% of Americans believe that people who are gay should be able to serve in the US military, with a reported margin of error of 3%[8]. The margin of error tells us how far off the actual value could be from the survey value[9]. Express the set of possible values using absolute values.

Since we want the size of the difference between the actual percentage, p, and the reported percentage to be less than 3%,

$$|p-78| \leq 3$$

---

[8] http://www.pollingreport.com/civil.htm, retrieved August 4, 2010

[9] Technically, margin of error usually means that the surveyors are 95% confident that actual value falls within this range.

Try it Now

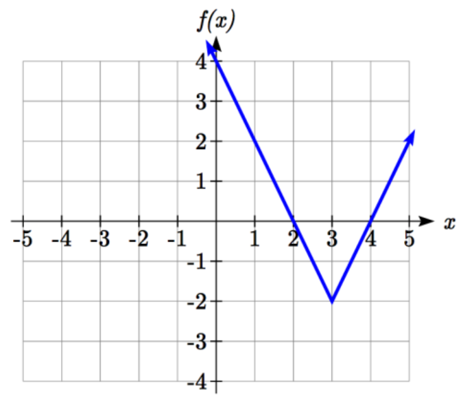
1. Students who score within 20 points of 80 will pass the test. Write this as a distance from 80 using the absolute value notation.

Important Features

The most significant feature of the absolute value graph is the corner point where the graph changes direction. When finding the equation for a transformed absolute value function, this point is very helpful for determining the horizontal and vertical shifts.

Example 3

Write an equation for the function graphed.



The basic absolute value function changes direction at the origin, so this graph has been shifted to the right 3 and down 2 from the basic toolkit function.

We might also notice that the graph appears stretched, since the linear portions have slopes of 2 and -2. From this information we could write the write the equation in two ways:

$f(x) = 2|x-3|-2$, treating the stretch as a vertical stretch

$f(x) = |2(x-3)|-2$, treating the stretch as a horizontal compression

Note that these equations are algebraically equivalent – the stretch for an absolute value function can be written interchangeably as a vertical or horizontal stretch/compression.

If you had not been able to determine the stretch based on the slopes of the lines, you can solve for the stretch factor by putting in a known pair of values for *x* and *f(x)*

$f(x) = a|x-3|-2$ Now substituting in the point (1, 2)

$2 = a|1-3|-2$

$4 = 2a$

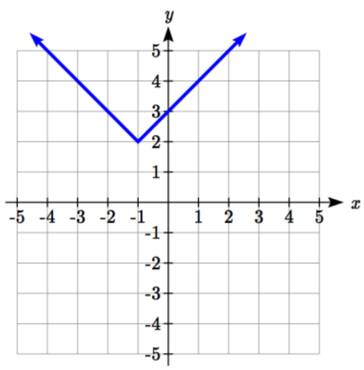
$a = 2$

Try it Now

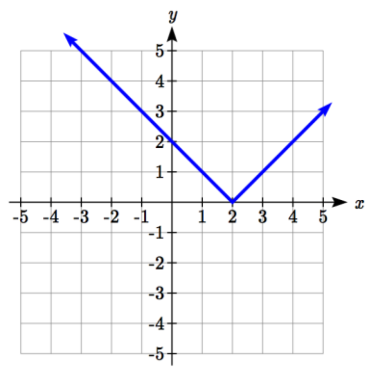
2. Given the description of the transformed absolute value function write the equation. The absolute value function is horizontally shifted left 2 units, is vertically flipped, and vertically shifted up 3 units.

The graph of an absolute value function will have a vertical intercept when the input is zero. The graph may or may not have horizontal intercepts, depending on how the graph has been shifted and reflected. It is possible for the absolute value function to have zero, one, or two horizontal intercepts.

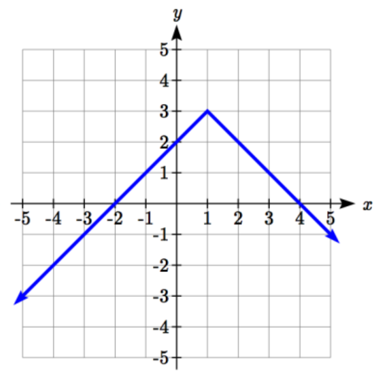
Zero horizontal intercepts



One



Two



To find the horizontal intercepts, we will need to solve an equation involving an absolute value.

Notice that the absolute value function is not one-to-one, so typically inverses of absolute value functions are not discussed.

Solving Absolute Value Equations

To solve an equation like $8 = |2x - 6|$, we can notice that the absolute value will be equal to eight if the quantity *inside* the absolute value were 8 or -8. This leads to two different equations we can solve independently:

$2x - 6 = 8$ or $2x - 6 = -8$
$2x = 14$ $\quad$ $2x = -2$
$x = 7$ $\quad$ $x = -1$

Solutions to Absolute Value Equations

An equation of the form $|A| = B$, with $B \geq 0$, will have solutions when

$A = B$ or $A = -B$

Example 4

Find the horizontal intercepts of the graph of $f(x) = |4x + 1| - 7$

The horizontal intercepts will occur when $f(x) = 0$. Solving,

$0 = |4x+1| - 7$ Isolate the absolute value on one side of the equation

$7 = |4x+1|$ Now we can break this into two separate equations:

$7 = 4x + 1$ or $-7 = 4x + 1$

$6 = 4x$ $\quad$ $-8 = 4x$

$x = \frac{6}{4} = \frac{3}{2}$ $\quad$ $x = \frac{-8}{4} = -2$

The graph has two horizontal intercepts, at $x = \frac{3}{2}$ and $x = -2$

Example 5

Solve $1 = 4|x-2| + 2$

Isolating the absolute value on one side the equation,

$1 = 4|x-2| + 2$

$-1 = 4|x-2|$

$-\frac{1}{4} = |x-2|$

At this point, we notice that this equation has no solutions – the absolute value always returns a positive value, so it is impossible for the absolute value to equal a negative value.

Try it Now

3. Find the horizontal & vertical intercepts for the function $f(x) = -|x+2| + 3$

Solving Absolute Value Inequalities

When absolute value inequalities are written to describe a set of values, like the inequality $|x-5| \le 4$ we wrote earlier, it is sometimes desirable to express this set of values without the absolute value, either using inequalities, or using interval notation.

We will explore two approaches to solving absolute value inequalities:

1) Using the graph
2) Using test values

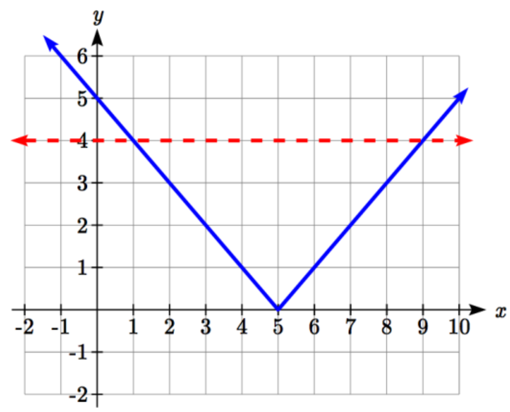
Example 6

Solve $|x-5| \le 4$

With both approaches, we will need to know first where the corresponding *equality* is true. In this case, we first will find where $|x-5| = 4$. We do this because the absolute value is a nice friendly function with no breaks, so the only way the function values can switch from being less than 4 to being greater than 4 is by passing through where the values equal 4. Solve $|x-5| = 4$,

$$x-5=4 \qquad \text{or} \qquad x-5=-4$$
$$x=9 \qquad\qquad\qquad x=1$$

To use a graph, we can sketch the function $f(x)=|x-5|$. To help us see where the outputs are 4, the line $g(x)=4$ could also be sketched.



On the graph, we can see that indeed the output values of the absolute value are equal to 4 at $x = 1$ and $x = 9$. Based on the shape of the graph, we can determine the absolute value is less than or equal to 4 between these two points, when $1 \le x \le 9$. In interval notation, this would be the interval [1,9].

As an alternative to graphing, after determining that the absolute value is equal to 4 at $x = 1$ and $x = 9$, we know the graph can only change from being less than 4 to greater than 4 at these values. This divides the number line up into three intervals: $x<1$, $1<x<9$, and $x>9$. To determine when the function is less than 4, we could pick a value in each interval and see if the output is less than or greater than 4.

Interval	Test x	*f(x)*	<4 or >4?
$x<1$	0	$\|0-5\|=5$	greater
$1<x<9$	6	$\|6-5\|=1$	less
$x>9$	11	$\|11-5\|=6$	greater

Since $1 \le x \le 9$ is the only interval in which the output at the test value is less than 4, we can conclude the solution to $|x-5| \le 4$ is $1 \le x \le 9$.

Example 7

Given the function $f(x) = -\frac{1}{2}|4x-5|+3$, determine for what x values the function values are negative.

We are trying to determine where *f(x)* < 0, which is when $-\frac{1}{2}|4x-5|+3<0$. We begin by isolating the absolute value:

$-\frac{1}{2}|4x-5|<-3$ when we multiply both sides by -2, it reverses the inequality

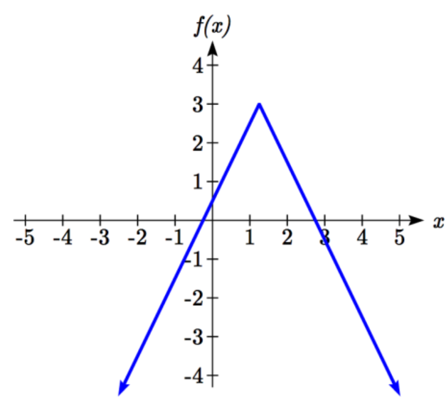
$|4x-5|>6$

Next we solve for the equality $|4x-5|=6$

$4x-5=6$		$4x-5=-6$
$4x=11$	or	$4x=-1$
$x=\frac{11}{4}$		$x=\frac{-1}{4}$

We can now either pick test values or sketch a graph of the function to determine on which intervals the original function value are negative. Notice that it is not even really important exactly what the graph looks like, as long as we know that it crosses the horizontal axis at $x=\frac{-1}{4}$ and $x=\frac{11}{4}$, and that the graph has been reflected vertically.

From the graph of the function, we can see the function values are negative to the left of the first horizontal intercept at $x=\frac{-1}{4}$, and negative to the right of the second intercept at $x=\frac{11}{4}$. This gives us the solution to the inequality:



$$x<\frac{-1}{4} \quad or \quad x>\frac{11}{4}$$

In interval notation, this would be $\left(-\infty,\frac{-1}{4}\right)\cup\left(\frac{11}{4},\infty\right)$

Try it Now

4. Solve $-2|k-4| \le -6$

Important Topics of this Section
The properties of the absolute value function
Solving absolute value equations
Finding intercepts
Solving absolute value inequalities

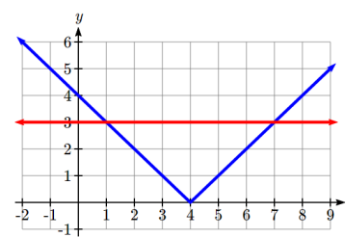
Try it Now Answers

1. Using the variable *p*, for passing, $|p-80| \le 20$

2. $f(x) = -|x+2|+3$

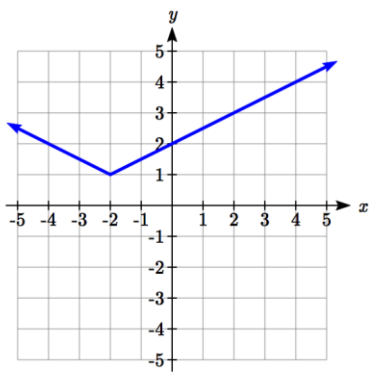
3. $f(0) = 1$, so the vertical intercept is at (0,1).
   $f(x) = 0$ when
   $-|x+2|+3=0$
   $|x+2|=3$
   $x+2=3$ or $x+2=-3$
   $x = 1$ or $x = -5$ so the horizontal intercepts are at (-5,0) & (1,0)

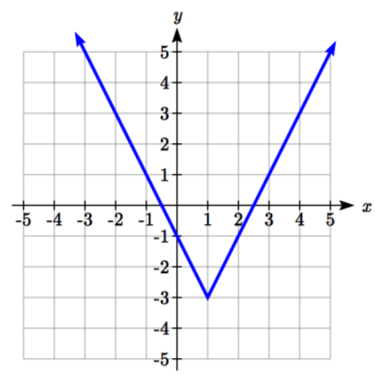
4. $-2|k-4| \le -6$
   $|k-4| \ge 3$
   Solving the equality $|k-4|=3$, $k-4=3$ or $k-4=-3$, so $k=1$ or $k=7$.
   Using a graph or test values, we can determine the intervals that satisfy the inequality are $k \le 1$ or $k \ge 7$; in interval notation this would be $(-\infty,1] \cup [7,\infty)$

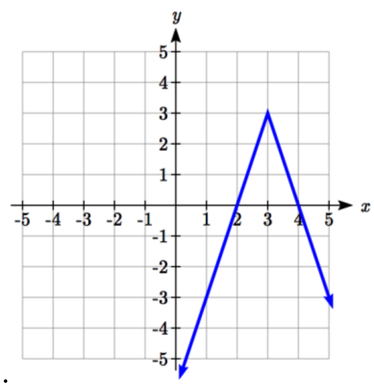


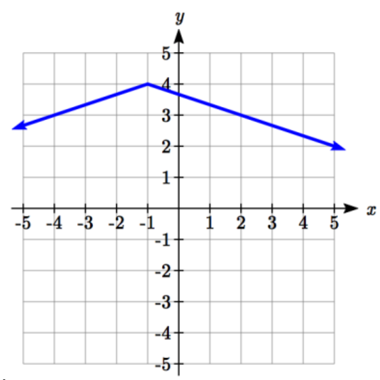
Section 2.5 Exercises

Write an equation for each transformation of $f(x) = |x|$

1. 

2. 

3. 

4. 

Sketch a graph of each function

5. $f(x) = -|x-1| - 1$
6. $f(x) = -|x+3| + 4$
7. $f(x) = 2|x+3| + 1$
8. $f(x) = 3|x-2| - 3$
9. $f(x) = |2x-4| - 3$
10. $f(x) = |3x+9| + 2$

Solve each the equation

11. $|5x-2| = 11$
12. $|4x+2| = 15$
13. $2|4-x| = 7$
14. $3|5-x| = 5$
15. $3|x+1| - 4 = -2$
16. $5|x-4| - 7 = 2$

Find the horizontal and vertical intercepts of each function

17. $f(x) = 2|x+1| - 10$

18. $f(x) = 4|x-3| + 4$

19. $f(x) = -3|x-2| - 1$

20. $f(x) = -2|x+1| + 6$

Solve each inequality

21. $|x+5| < 6$

22. $|x-3| < 7$

23. $|x-2| \geq 3$

24. $|x+4| \geq 2$

25. $|3x+9| < 4$

26. $|2x-9| \leq 8$

4. Describing Bivariate Data

A dataset with two variables contains what is called bivariate data. This chapter discusses ways to describe the relationship between two variables. For example, you may wish to describe the relationship between the heights and weights of people to determine the extent to which taller people weigh more.

The introductory section gives more examples of bivariate relationships and presents the most common way of portraying these relationships graphically. The next five sections discuss Pearson's correlation, the most common index of the relationship between two variables. The final section, "Variance Sum Law II," makes use of Pearson's correlation to generalize this law to bivariate data.

Introduction to Bivariate Data

by Rudy Guerra and David M. Lane

Prerequisites
- Chapter 1: Variables
- Chapter 1: Distributions
- Chapter 2: Histograms
- Chapter 3: Measures of Central Tendency
- Chapter 3: Variability
- Chapter 3: Shapes of Distributions

Learning Objectives
1. Define "bivariate data"
2. Define "scatter plot"
3. Distinguish between a linear and a nonlinear relationship
4. Identify positive and negative associations from a scatter plot

Measures of central tendency, variability, and spread summarize a single variable by providing important information about its distribution. Often, more than one variable is collected on each individual. For example, in large health studies of populations it is common to obtain variables such as age, sex, height, weight, blood pressure, and total cholesterol on each individual. Economic studies may be interested in, among other things, personal income and years of education. As a third example, most university admissions committees ask for an applicant's high school grade point average and standardized admission test scores (e.g., SAT). In this chapter we consider bivariate data, which for now consists of two quantitative variables for each individual. Our first interest is in summarizing such data in a way that is analogous to summarizing univariate (single variable) data.

By way of illustration, let's consider something with which we are all familiar: age. Let's begin by asking if people tend to marry other people of about the same age. Our experience tells us "yes," but how good is the correspondence? One way to address the question is to look at pairs of ages for a sample of married couples. Table 1 below shows the ages of 10 married couples. Going across the columns we see that, yes, husbands and wives tend to be of about the same age, with men having a tendency to be slightly older than their wives. This is no big surprise, but at least the data bear out our experiences, which is not always the case.

Table 1. Sample of spousal ages of 10 White American Couples.

Husband	36	72	37	36	51	50	47	50	37	41
Wife	35	67	33	35	50	46	47	42	36	41

The pairs of ages in Table 1 are from a dataset consisting of 282 pairs of spousal ages, too many to make sense of from a table. What we need is a way to summarize the 282 pairs of ages. We know that each variable can be summarized by a histogram (see Figure 1) and by a mean and standard deviation (See Table 2).

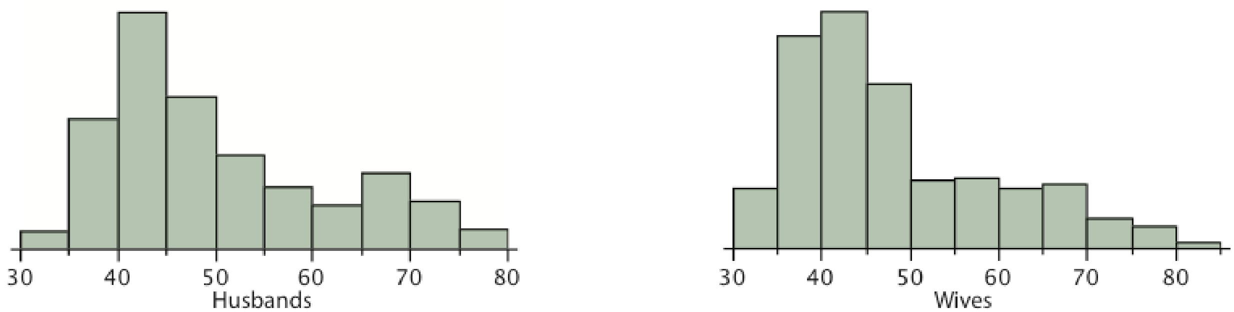


Figure 1. Histograms of spousal ages.

Table 2. Means and standard deviations of spousal ages.

	Mean	**Standard Deviation**
Husbands	49	11
Wives	47	11

Each distribution is fairly skewed with a long right tail. From Table 1 we see that not all husbands are older than their wives and it is important to see that this fact is lost when we separate the variables. That is, even though we provide summary statistics on each variable, the pairing within couple is lost by separating the variables. We cannot say, for example, based on the means alone what percentage of couples has younger husbands than wives. We have to count across pairs to find this out. Only by maintaining the pairing can meaningful answers be found about couples per se. Another example of information not available from the separate descriptions of husbands and wives' ages is the mean age of husbands with wives of a certain age. For instance, what is the average age of husbands with 45-year-old wives? Finally, we do not know the relationship between the husband's age and the wife's age.

We can learn much more by displaying the bivariate data in a graphical form that maintains the pairing. Figure 2 shows a scatter plot of the paired ages. The x-axis represents the age of the husband and the y-axis the age of the wife.

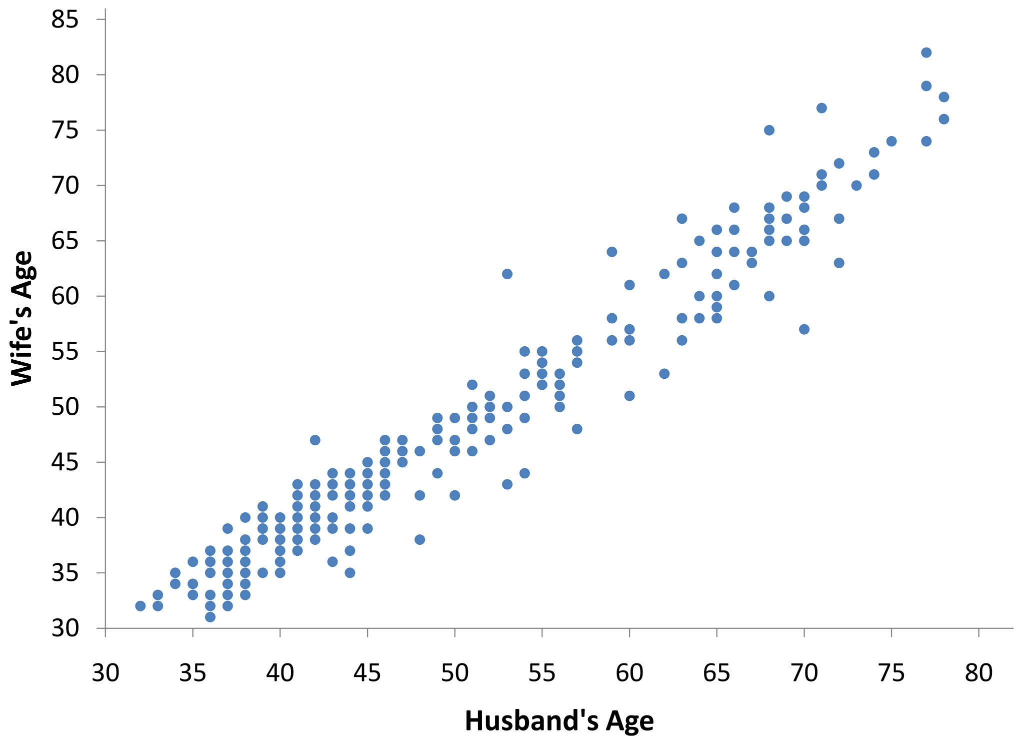


Figure 2. Scatter plot showing wife’s age as a function of husband’s age.

There are two important characteristics of the data revealed by Figure 2. First, it is clear that there is a strong relationship between the husband's age and the wife's age: the older the husband, the older the wife. When one variable (Y) increases with the second variable (X), we say that X and Y have a positive association. Conversely, when Y decreases as X increases, we say that they have a negative association.

Second, the points cluster along a straight line. When this occurs, the relationship is called a linear relationship.

Figure 3 shows a scatter plot of Arm Strength and Grip Strength from 149 individuals working in physically demanding jobs including electricians, construction and maintenance workers, and auto mechanics. Not surprisingly, the stronger someone's grip, the stronger their arm tends to be. There is therefore a positive association between these variables. Although the points cluster along a line, they are not clustered quite as closely as they are for the scatter plot of spousal age.

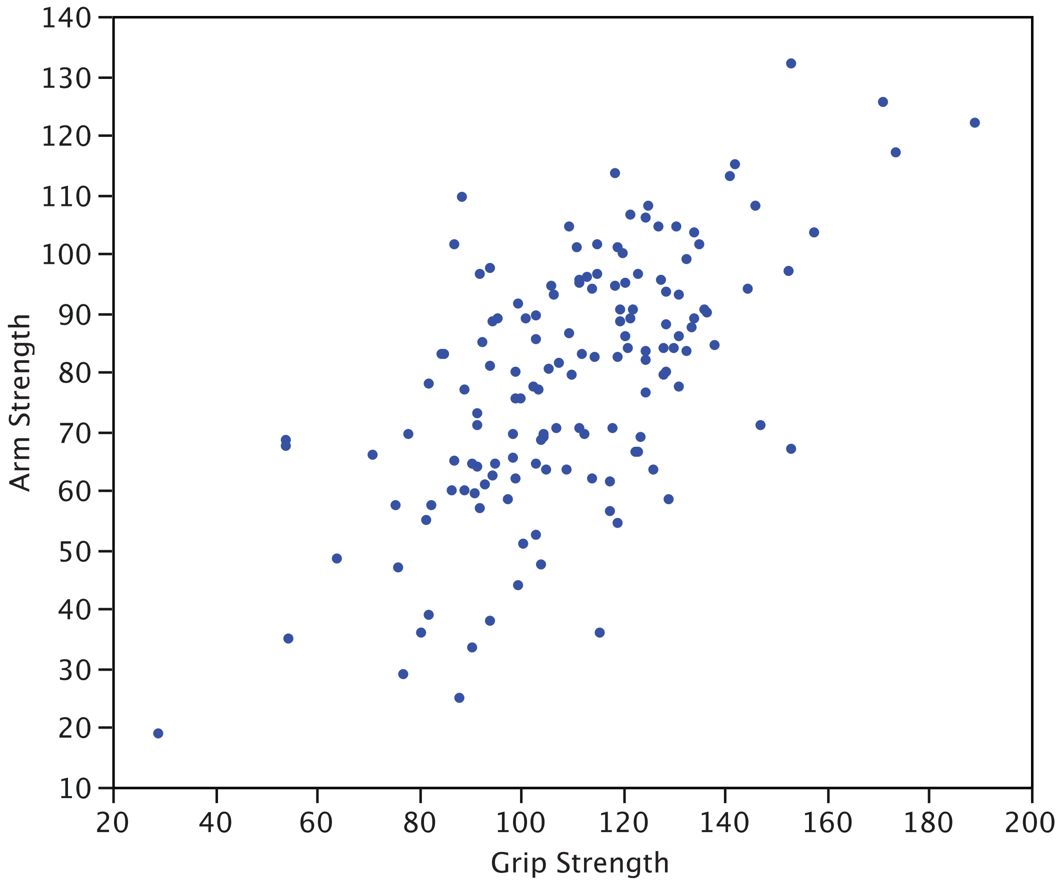


Figure 3. Scatter plot of Grip Strength and Arm Strength.

Not all scatter plots show linear relationships. Figure 4 shows the results of an experiment conducted by Galileo on projectile motion. In the experiment, Galileo rolled balls down an incline and measured how far they traveled as a function of the release height. It is clear from Figure 4 that the relationship between “Release Height” and “Distance Traveled” is not described well by a straight line: If you drew a line connecting the lowest point and the highest point, all of the remaining points would be above the line. The data are better fit by a parabola.

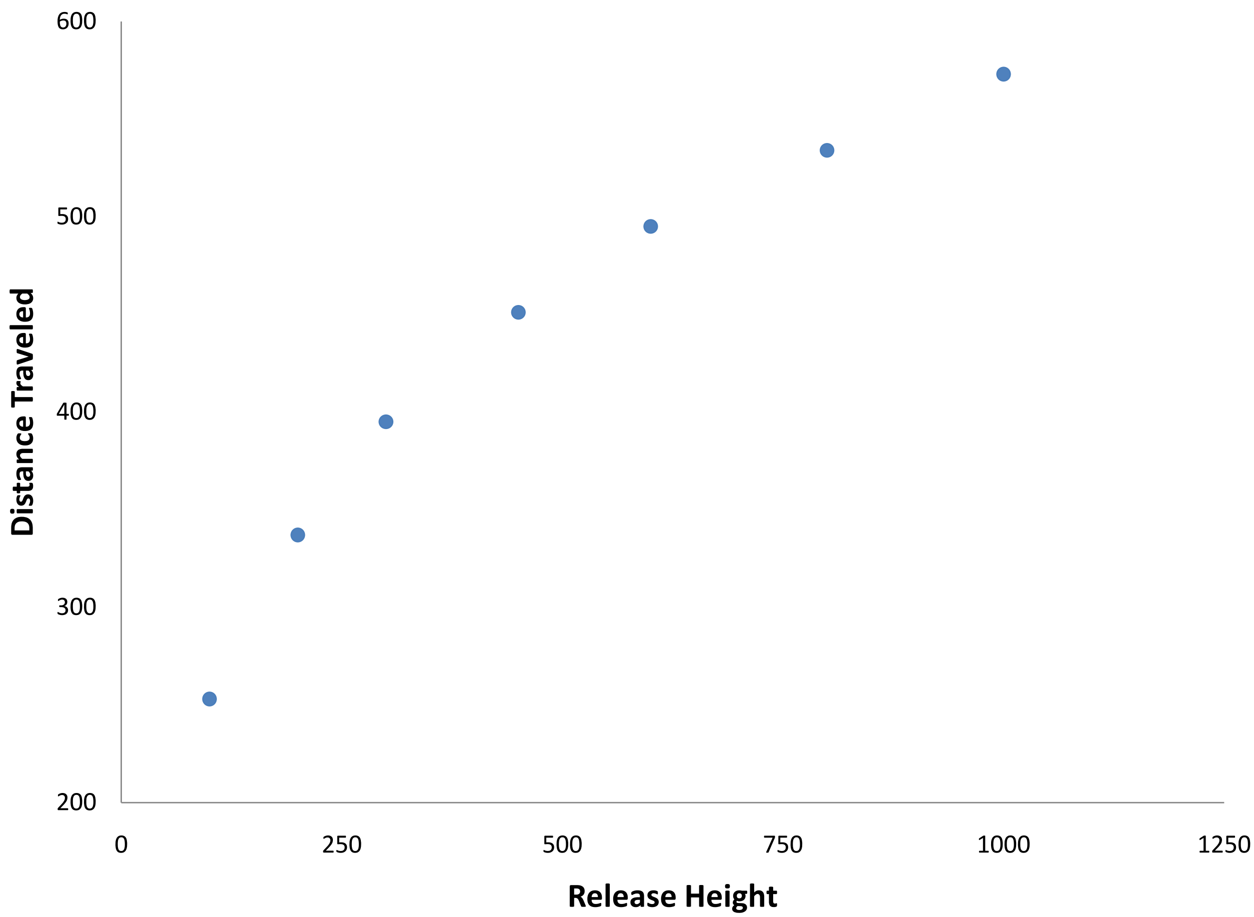


Figure 4. Galileo's data showing a non-linear relationship.

Scatter plots that show linear relationships between variables can differ in several ways including the slope of the line about which they cluster and how tightly the points cluster about the line. A statistical measure of the strength of the relationship between two quantitative variables that takes these factors into account is the subject of the next section.

Values of the Pearson Correlation

by David M. Lane

Prerequisites
• Chapter 4: Introduction to Bivariate Data

Learning Objectives
1. Describe what Pearson's correlation measures
2. Give the symbols for Pearson's correlation in the sample and in the population
3. State the possible range for Pearson's correlation
4. Identify a perfect linear relationship

The Pearson product-moment correlation coefficient is a measure of the strength of the linear relationship between two variables. It is referred to as Pearson's correlation or simply as the correlation coefficient. If the relationship between the variables is not linear, then the correlation coefficient does not adequately represent the strength of the relationship between the variables.

The symbol for Pearson's correlation is "ρ" when it is measured in the population and "r" when it is measured in a sample. Because we will be dealing almost exclusively with samples, we will use r to represent Pearson's correlation unless otherwise noted.

Pearson's r can range from -1 to 1. An r of -1 indicates a perfect negative linear relationship between variables, an r of 0 indicates no linear relationship between variables, and an r of 1 indicates a perfect positive linear relationship between variables. Figure 1 shows a scatter plot for which r = 1.

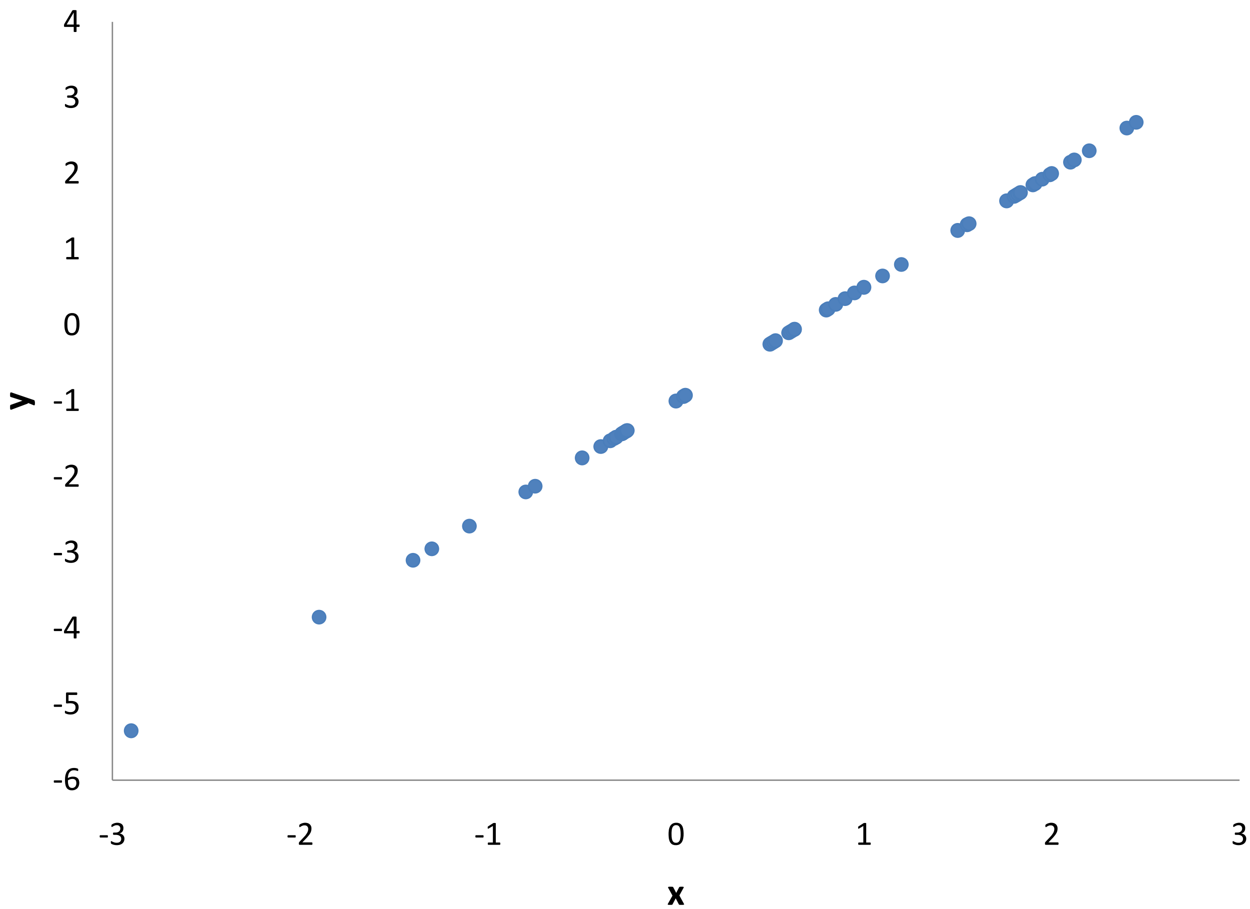


Figure 1. A perfect linear relationship, r = 1.

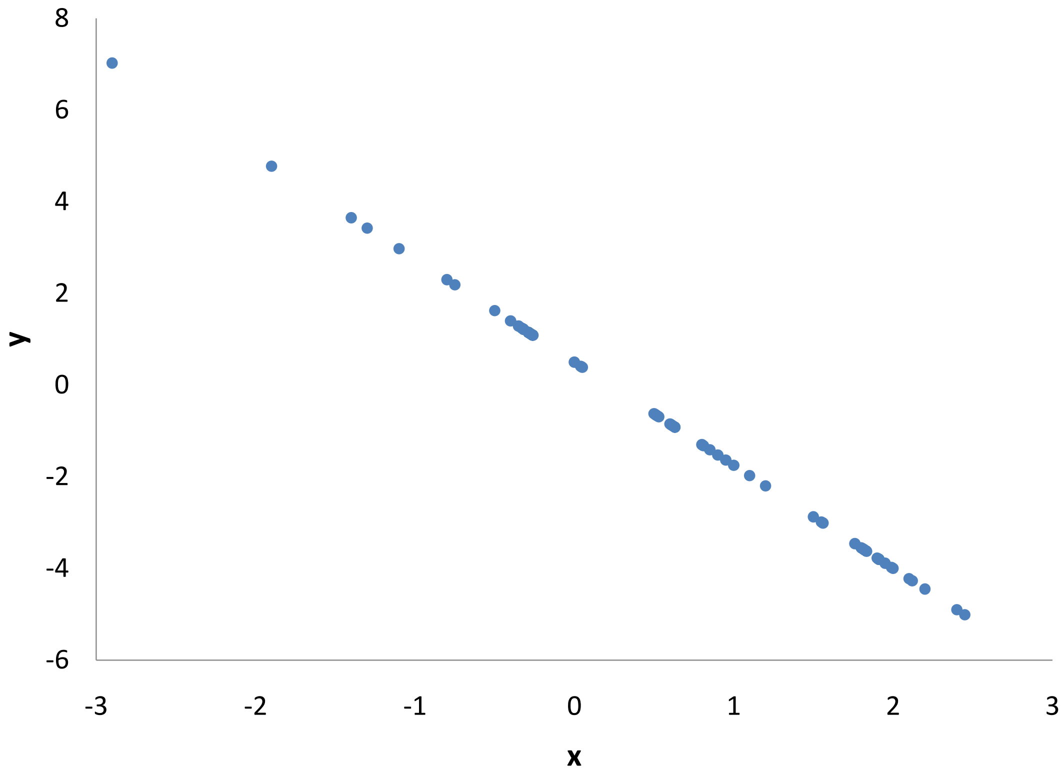


Figure 2. A perfect negative linear relationship, r = -1.

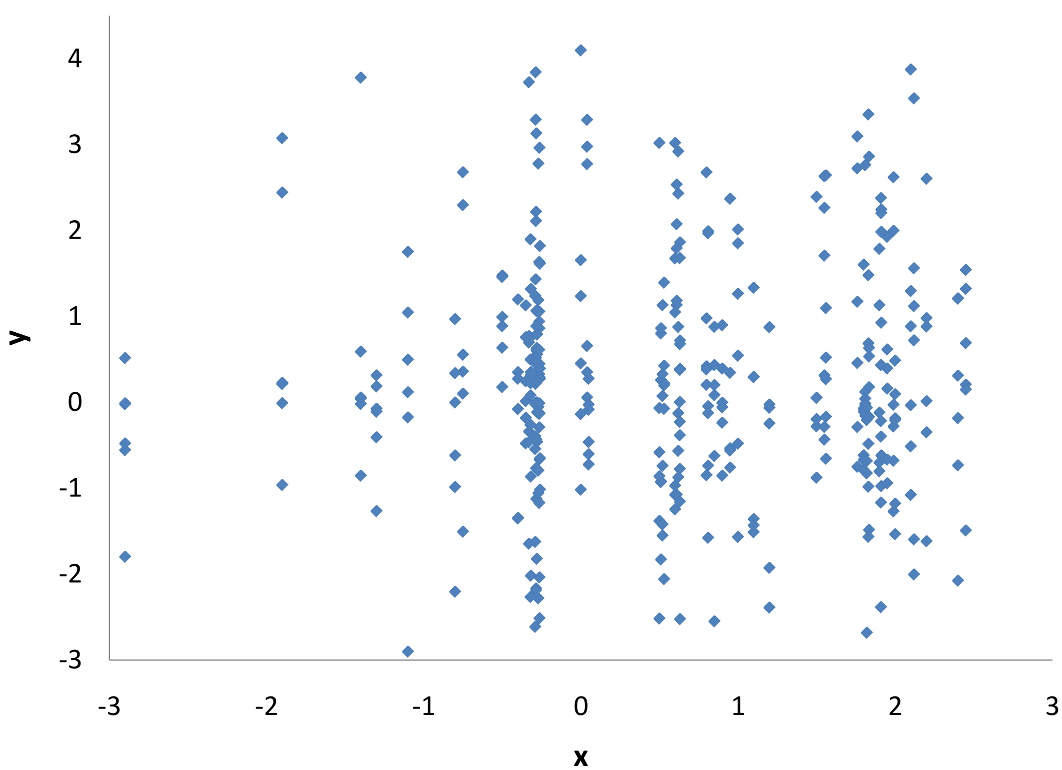


Figure 3. A scatter plot for which r = 0. Notice that there is no relationship between X and Y.

With real data, you would not expect to get values of r of exactly -1, 0, or 1. The data for spousal ages shown in Figure 4 and described in the introductory section has an r of 0.97.

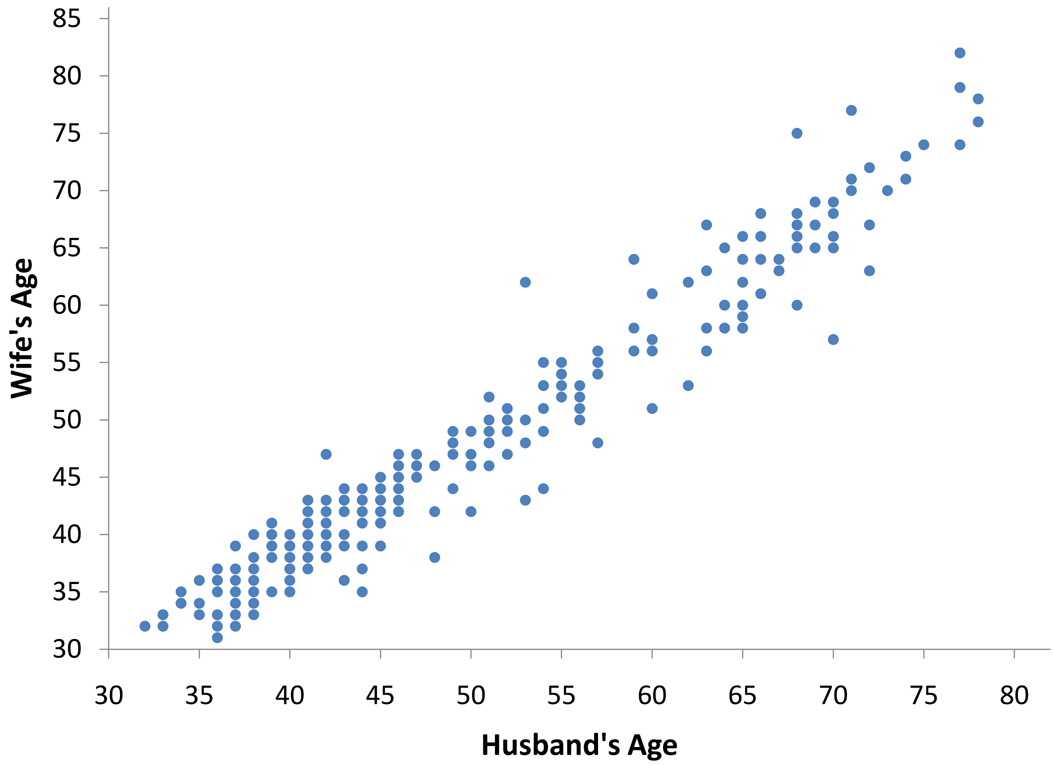


Figure 4. Scatter plot of spousal ages, r = 0.97.

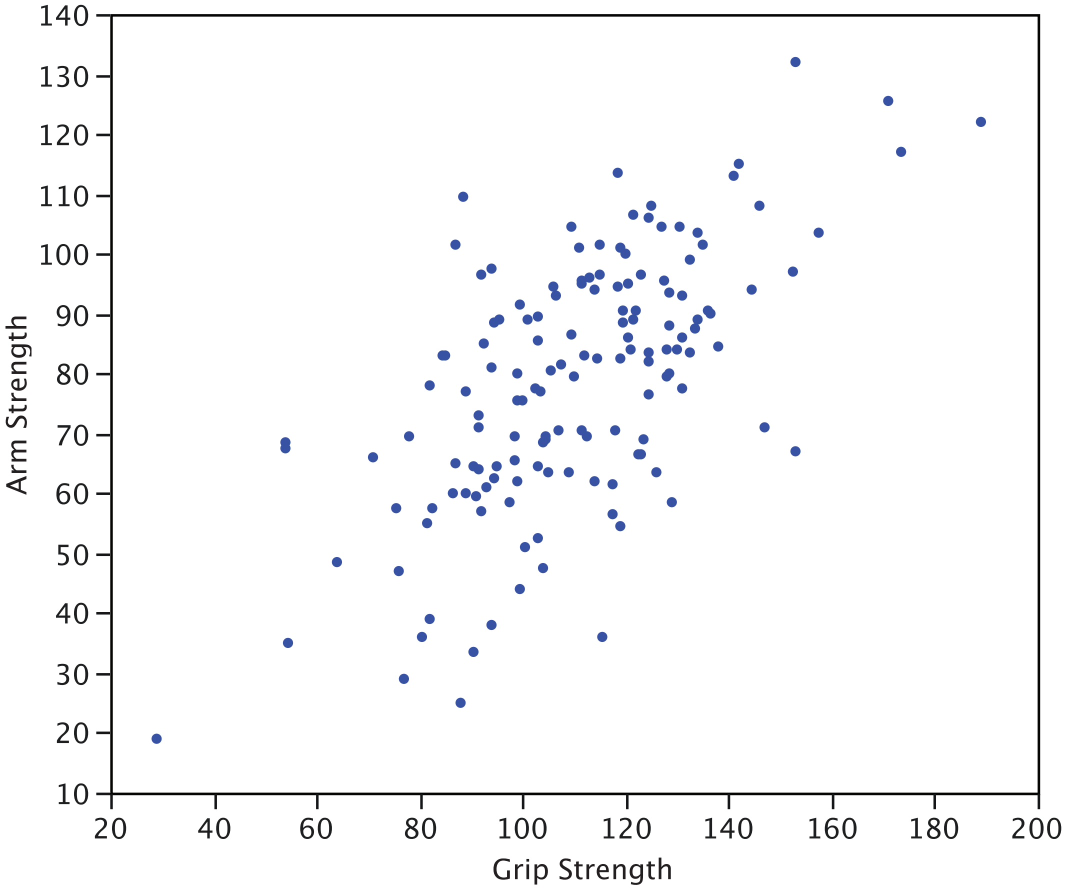


Figure 5. Scatter plot of Grip Strength and Arm Strength, r = 0.63.

The relationship between grip strength and arm strength depicted in Figure 5 (also described in the introductory section) is 0.63.

Properties of Pearson's r

by David M. Lane

Prerequisites
• Chapter 1: Linear Transformations
• Chapter 4: Introduction to Bivariate Data

Learning Objectives
1. State the range of values for Pearson's correlation
2. State the values that represent perfect linear relationships
3. State the relationship between the correlation of Y with X and the correlation of X with Y
4. State the effect of linear transformations on Pearson's correlation

A basic property of Pearson's r is that its possible range is from -1 to 1. A correlation of -1 means a perfect negative linear relationship, a correlation of 0 means no linear relationship, and a correlation of 1 means a perfect positive linear relationship.

Pearson's correlation is symmetric in the sense that the correlation of X with Y is the same as the correlation of Y with X. For example, the correlation of Weight with Height is the same as the correlation of Height with Weight.

A critical property of Pearson's r is that it is unaffected by linear transformations. This means that multiplying a variable by a constant and/or adding a constant does not change the correlation of that variable with other variables. For instance, the correlation of Weight and Height does not depend on whether Height is measured in inches, feet, or even miles. Similarly, adding five points to every student's test score would not change the correlation of the test score with other variables such as GPA.

Computing Pearson's r

by David M. Lane

Prerequisites
• Chapter 1: Summation Notation
• Chapter 4: Introduction to Bivariate Data

Learning Objectives
1. Define X and x
2. State why $\Sigma xy = 0$ when there is no relationship
3. Calculate r

There are several formulas that can be used to compute Pearson's correlation. Some formulas make more conceptual sense whereas others are easier to actually compute. We are going to begin with a formula that makes more conceptual sense.

We are going to compute the correlation between the variables X and Y shown in Table 1. We begin by computing the mean for X and subtracting this mean from all values of X. The new variable is called "x." The variable "y" is computed similarly. The variables x and y are said to be deviation scores because each score is a deviation from the mean. Notice that the means of x and y are both 0. Next we create a new column by multiplying x and y.

Before proceeding with the calculations, let's consider why the sum of the xy column reveals the relationship between X and Y. If there were no relationship between X and Y, then positive values of x would be just as likely to be paired with negative values of y as with positive values. This would make negative values of xy as likely as positive values and the sum would be small. On the other hand, consider Table 1 in which high values of X are associated with high values of Y and low values of X are associated with low values of Y. You can see that positive values of x are associated with positive values of y and negative values of x are associated with negative values of y. In all cases, the product of x and y is positive, resulting in a high total for the xy column. Finally, if there were a negative relationship then positive values of x would be associated with negative values of y and negative values of x would be associated with positive values of y. This would lead to negative values for xy.

Table 1. Calculation of r.

	X	Y	x	y	xy	x^2	y^2
	1	4	-3	-5	15	9	25
	3	6	-1	-3	3	1	9
	5	10	1	1	1	1	1
	5	12	1	3	3	1	9
	6	13	2	4	8	4	16
Total	20	45	0	0	30	16	60
Mean	4	9	0	0	6		

Pearson's r is designed so that the correlation between height and weight is the same whether height is measured in inches or in feet. To achieve this property, Pearson's correlation is computed by dividing the sum of the xy column (Σxy) by the square root of the product of the sum of the x^2 column (Σx^2) and the sum of the y^2 column (Σy^2). The resulting formula is:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

and therefore

$$r = \frac{30}{\sqrt{(16)(60)}} = \frac{30}{\sqrt{960}} = \frac{30}{30.984} = 0.968$$

An alternative computational formula that avoids the step of computing deviation scores is:

$$r = \frac{\sum xy - \frac{\sum x \sum y}{N}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{N}\right)}\sqrt{\left(\sum y^2 - \frac{(\sum y)^2}{N}\right)}}$$

Variance Sum Law II

by David M. Lane

Prerequisites
- Chapter 1: Variance Sum Law I
- Chapter 4: Values of Pearson's Correlation

Learning Objectives
1. State the variance sum law when X and Y are not assumed to be independent
2. Compute the variance of the sum of two variables if the variance of each and their correlation is known
3. Compute the variance of the difference between two variables if the variance of each and their correlation is known

Recall that when the variables X and Y are independent, the variance of the sum or difference between X and Y can be written as follows:

$$\sigma^2_{X \pm Y} = \sigma^2_X + \sigma^2_Y$$

which is read: "The variance of X plus or minus Y is equal to the variance of X plus the variance of Y."

When X and Y are correlated, the following formula should be used:

$$\sigma^2_{X \pm Y} = \sigma^2_X + \sigma^2_Y \pm 2\rho\sigma_X\sigma_Y$$

where ϱ is the correlation between X and Y in the population. For example, if the variance of verbal SAT were 10,000, the variance of quantitative SAT were 11,000 and the correlation between these two tests were 0.50, then the variance of total SAT (verbal + quantitative) would be:

$$\sigma^2_{verbal+quant} = 10{,}000 + 11{,}000 + (2)(0.5)\sqrt{10{,}000}\sqrt{11{,}000}$$

which is equal to 31,488. The variance of the difference is:

$$\sigma^2_{verbal-quant} = 10{,}000 + 11{,}000 - (2)(0.5)\sqrt{10{,}000}\sqrt{11{,}000}$$

which is equal to 10,512.

If the variances and the correlation are computed in a sample, then the following notation is used to express the variance sum law:

$$s_{X \pm Y}^2 = s_X^2 + s_Y^2 \pm 2rs_Xs_y$$

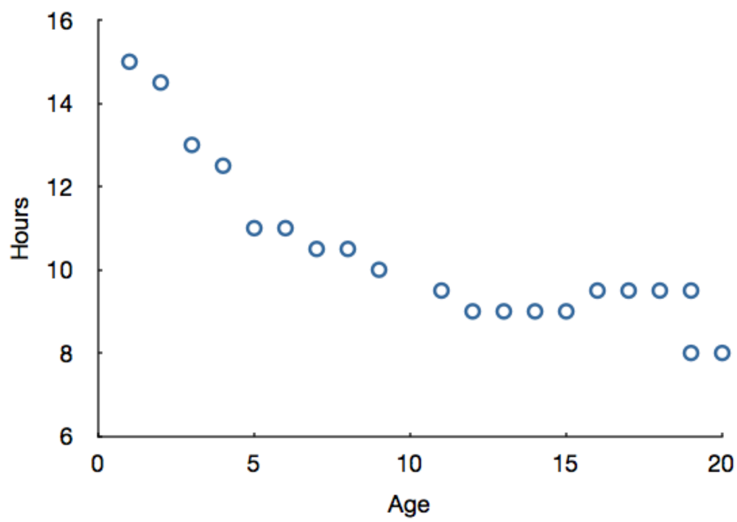
Statistical Literacy

by David M. Lane

Prerequisites

- Chapter 4: Values of Pearson's Correlation

The graph below showing the relationship between age and sleep is based on a graph that appears on [this web page].



What do you think?

Why might Pearson's correlation not be a good way to describe the relationship?

> Pearson's correlation measures the strength of the linear relationship between two variables. The relationship here is not linear. As age increases, hours slept decreases rapidly at first but then levels off.

14. Regression

This chapter is about prediction. Statisticians are often called upon to develop methods to predict one variable from other variables. For example, one might want to predict college grade point average from high school grade point average. Or, one might want to predict income from the number of years of education.

Introduction to Linear Regression

by David M. Lane

Prerequisites
• Chapter 3: Measures of Variability
• Chapter 4: Describing Bivariate Data

Learning Objectives
1. Define linear regression
2. Identify errors of prediction in a scatter plot with a regression line

In simple linear regression, we predict scores on one variable from the scores on a second variable. The variable we are predicting is called the *criterion variable* and is referred to as Y. The variable we are basing our predictions on is called the *predictor variable* and is referred to as X. When there is only one predictor variable, the prediction method is called *simple regression*. In simple linear regression, the topic of this section, the predictions of Y when plotted as a function of X form a straight line.

The example data in Table 1 are plotted in Figure 1. You can see that there is a positive relationship between X and Y. If you were going to predict Y from X, the higher the value of X, the higher your prediction of Y.

Table 1. Example data.

X	Y
1.00	1.00
2.00	2.00
3.00	1.30
4.00	3.75
5.00	2.25

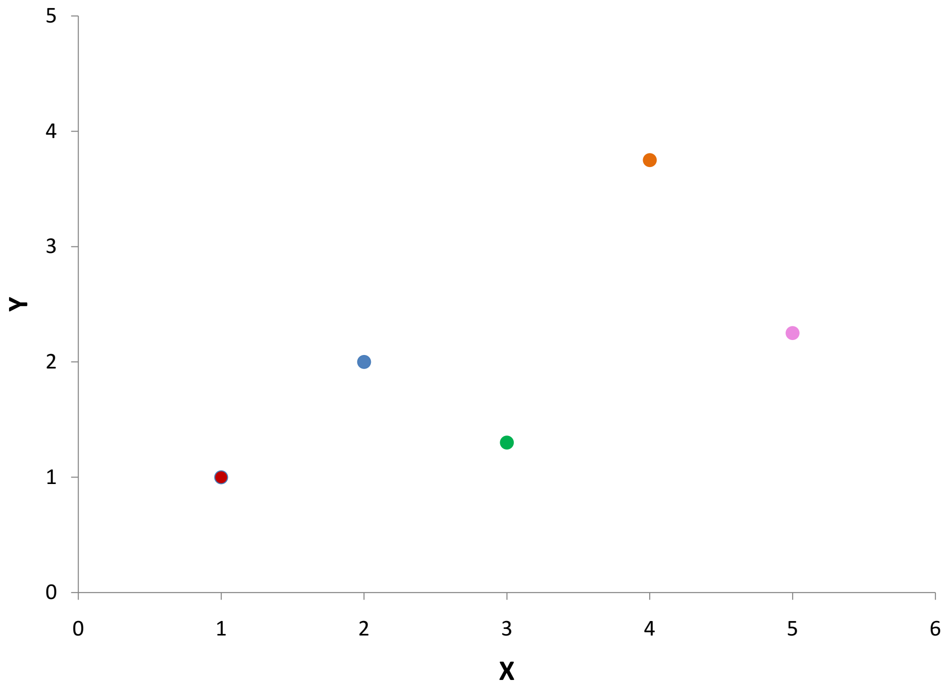


Figure 1. A scatter plot of the example data.

Linear regression consists of finding the best-fitting straight line through the points. The best-fitting line is called a regression line. The black diagonal line in Figure 2 is the regression line and consists of the predicted score on Y for each possible value of X. The vertical lines from the points to the regression line represent the errors of prediction. As you can see, the red point is very near the regression line; its error of prediction is small. By contrast, the yellow point is much higher than the regression line and therefore its error of prediction is large.

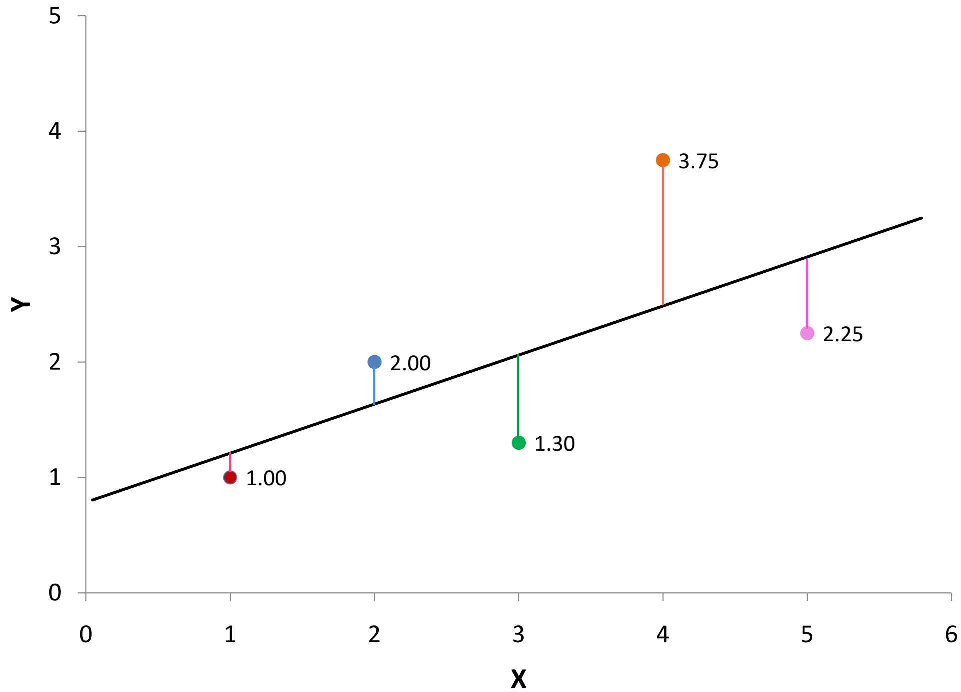


Figure 2. A scatter plot of the example data. The black line consists of the predictions, the points are the actual data, and the vertical lines between the points and the black line represent errors of prediction.

The error of prediction for a point is the value of the point minus the predicted value (the value on the line). Table 2 shows the predicted values (Y') and the errors of prediction (Y-Y'). For example, the first point has a Y of 1.00 and a predicted Y of 1.21. Therefore, its error of prediction is -0.21.

Table 2. Example data.

X	Y	Y'	Y-Y'	$(Y-Y')^2$
1.00	1.00	1.210	-0.210	0.044
2.00	2.00	1.635	0.365	0.133
3.00	1.30	2.060	-0.760	0.578
4.00	3.75	2.485	1.265	1.600
5.00	2.25	2.910	-0.660	0.436

You may have noticed that we did not specify what is meant by “best-fitting line.” By far the most commonly used criterion for the best-fitting line is the line that minimizes the sum of the squared errors of prediction. That is the criterion that was used to find the line in Figure 2. The last column in Table 2 shows the squared errors of prediction. The sum of the squared errors of prediction shown in Table 2 is lower than it would be for any other regression line.

The formula for a regression line is

$$Y' = bX + A$$

where Y' is the predicted score, b is the slope of the line, and A is the Y intercept. The equation for the line in Figure 2 is

$$Y' = 0.425X + 0.785$$

For X = 1,

$$Y' = (0.425)(1) + 0.785 = 1.21.$$

For X = 2,

$$Y' = (0.425)(2) + 0.785 = 1.64.$$

Computing the Regression Line

In the age of computers, the regression line is typically computed with statistical software. However, the calculations are relatively easy are given here for anyone who is interested. The calculations are based on the statistics shown in Table 3. M_X is the mean of X, M_Y is the mean of Y, s_X is the standard deviation of X, s_Y is the standard deviation of Y, and r is the correlation between X and Y.

Table 3. Statistics for computing the regression line

M_X	M_Y	s_X	s_Y	r
3	2.06	1.581	1.072	0.627

The slope (b) can be calculated as follows:

$$b = r\frac{s_y}{s_x}$$

and the intercept (A) can be calculated as

$A = M_Y - bM_X$.

For these data,

$$b = (0.627)\frac{1.072}{1.581} = 0.425$$

$A = 2.06 - (0.425)(3) = 0.785$

Note that the calculations have all been shown in terms of sample statistics rather than population parameters. The formulas are the same; simply use the parameter values for means, standard deviations, and the correlation.

Standardized Variables

The regression equation is simpler if variables are *standardized* so that their means are equal to 0 and standard deviations are equal to 1, for then b = r and A = 0. This makes the regression line:

$Z_{Y'} = (r)(Z_X)$

where $Z_{Y'}$ is the predicted standard score for Y, r is the correlation, and Z_X is the standardized score for X. Note that the slope of the regression equation for standardized variables is r.

Figure 3 shows a scatterplot with the regression line predicting the standardized Verbal SAT from the standardized Math SAT.

A Real Example

The case study, “SAT and College GPA” contains high school and university grades for 105 computer science majors at a local state school. We now consider how we could predict a student's university GPA if we knew his or her high school GPA.

Figure 3 shows a scatter plot of University GPA as a function of High School GPA. You can see from the figure that there is a strong positive relationship. The correlation is 0.78. The regression equation is

Univ GPA' = (0.675)(High School GPA) + 1.097

Therefore, a student with a high school GPA of 3 would be predicted to have a university GPA of

```
University GPA' = (0.675)(3) + 1.097 = 3.12.
```

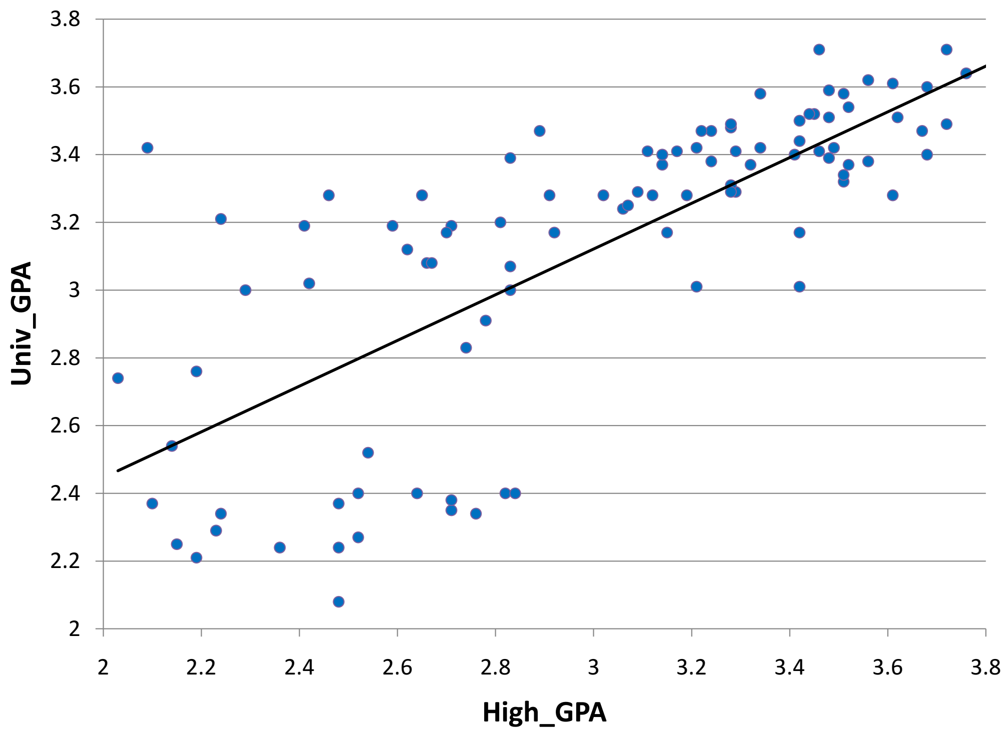


Figure 3. University GPA as a function of High School GPA.

Assumptions

It may surprise you, but the calculations shown in this section are assumption free. Of course, if the relationship between X and Y is not linear, a different shaped function could fit the data better. *Inferential statistics* in regression are based on several assumptions, and these assumptions are presented in a later section of this chapter.

Partitioning the Sums of Squares

by David M. Lane

Prerequisites

• Chapter 14: Introduction to Linear Regression

Learning Objectives

1. Compute the sum of squares Y
2. Convert raw scores to deviation scores
3. Compute predicted scores from a regression equation
4. Partition sum of squares Y into sum of squares predicted and sum of squares error
5. Define r^2 in terms of sum of squares explained and sum of squares Y

One useful aspect of regression is that it can divide the variation in Y into two parts: the variation of the predicted scores and the variation in the errors of prediction. The variation of Y is called the sum of squares Y and is defined as the sum of the squared deviations of Y from the mean of Y. In the population, the formula is

$$SSY = \sum (Y - \mu_Y)^2$$

where SSY is the sum of squares Y, Y is an individual value of Y, and my is the mean of Y. A simple example is given in Table 1. The mean of Y is 2.06 and SSY is the sum of the values in the third column and is equal to 4.597.

Table 1. Example of SSY.

Y	Y-m_y	(Y-m_y)^2
1.00	-1.06	1.1236
2.00	-0.06	0.0036
1.30	-0.76	0.5776
3.75	1.69	2.8561
2.25	0.19	0.0361

When computed in a sample, you should use the sample mean, M, in place of the population mean:

$$SSY = \sum (Y - M_Y)^2$$

It is sometimes convenient to use formulas that use *deviation scores* rather than raw scores. Deviation scores are simply deviations from the mean. By convention, small letters rather than capitals are used for deviation scores. Therefore, the score, y indicates the difference between Y and the mean of Y. Table 2 shows the use of this notation. The numbers are the same as in Table 1.

Table 2. Example of SSY using Deviation Scores.

Y	y	y^2
1.00	-1.06	1.1236
2.00	-0.06	0.0036
1.30	-0.76	0.5776
3.75	1.69	2.8561
2.25	0.19	0.0361
10.30	0.00	4.5970

The data in Table 3 are reproduced from the introductory section. The column X has the values of the *predictor variable* and the column Y has the *criterion variable*. The third column, y, contains the the differences between the column Y and the mean of Y.

Table 3. Example data. The last row contains column sums.

X	Y	y	y^2	Y'	y'	y'^2	Y-Y'	$(Y-Y')^2$
1.00	1.00	-1.06	1.1236	1.210	-0.850	0.7225	-0.210	0.044
2.00	2.00	-0.06	0.0036	1.635	-0.425	0.1806	0.365	0.133
3.00	1.30	-0.76	0.5776	2.060	0.000	0.000	-0.760	0.578
4.00	3.75	1.69	2.8561	2.485	0.425	0.1806	1.265	1.600
5.00	2.25	0.19	0.0361	2.910	0.850	0.7225	-0.660	0.436
15.00	10.30	0.00	4.597	10.300	0.000	1.806	0.000	2.791

The fourth column, y^2, is simply the square of the y column. The column Y' contains the predicted values of Y. In the introductory section, it was shown that the equation for the regression line for these data is

```
Y' = 0.425X + 0.785.
```

The values of Y' were computed according to this equation. The column y' contains deviations of Y' from the mean of Y' and y'^2 is the square of this column. The next-to-last column, Y-Y', contains the actual scores (Y) minus the predicted scores (Y'). The last column contains the squares of these errors of prediction.

We are now in a position to see how the SSY is partitioned. Recall that SSY is the sum of the squared deviations from the mean. It is therefore the sum of the y^2 column and is equal to 4.597. SSY can be partitioned into two parts: the sum of squares predicted (SSY') and the sum of squares error (SSE). The sum of squares predicted is the sum of the squared deviations of the predicted scores from the mean predicted score. In other words, it is the sum of the y'^2 column and is equal to 1.806. The sum of squares error is the sum of the squared errors of prediction. It is therefore the sum of the $(Y-Y')^2$ column and is equal to 2.791. This can be summed up as:

```
SSY = SSY' + SSE
4.597 = 1.806 + 2.791
```

There are several other notable features about Table 3. First, notice that the sum of y and the sum of y' are both zero. This will always be the case because these variables were created by subtracting their respective means from each value. Also, notice that the mean of Y-Y' is 0. This indicates that although some Y values are higher than their respective predicted Y values and some are lower, the average difference is zero.

The SSY is the total variation, the SSY' is the variation explained, and the SSE is the variation unexplained. Therefore, the proportion of variation explained can be computed as:

$$Proportion\ explained = \frac{SSY'}{SSY}$$

Similarly, the proportion not explained is:

$$Proportion\ not\ explained = \frac{SSE}{SSY}$$

There is an important relationship between the proportion of variation explained and Pearson's correlation: r^2 is the proportion of variation explained. Therefore, if r = 1, then, naturally, the proportion of variation explained is 1; if r = 0, then the proportion explained is 0. One last example: for r = 0.4, the proportion of variation explained is 0.16.

Since the variance is computed by dividing the variation by N (for a population) or N-1 (for a sample), the relationships spelled out above in terms of variation also hold for variance. For example,

$$\sigma^2_{total} = \sigma^2_{Y'} + \sigma^2_e$$

where the first term is the variance total, the second term is the variance of Y', and the last term is the variance of the errors of prediction (Y-Y'). Similarly, r^2 is the proportion of variance explained as well as the proportion of variation explained.

Summary Table

It is often convenient to summarize the partitioning of the data in a table such as Table 4. The *degrees of freedom* column (df) shows the degrees of freedom for each source of variation. The degrees of freedom for the sum of squares explained is equal to the number of predictor variables. This will always be 1 in simple regression. The error degrees of freedom is equal to the total number of observations minus 2. In this example, it is 5 - 2 = 3. The total degrees of freedom is the total number of observations minus 1.

Table 4. Summary Table for Example Data

Source	Sum of Squares	df	Mean Square
Explained	1.806	1	1.806
Error	2.791	3	0.930
Total	4.597	4	

Standard Error of the Estimate

by David M. Lane

Prerequisites
- Chapter 3: Measures of Variability
- Chapter 14: Introduction to Linear Regression
- Chapter 14: Partitioning Sums of Squares

Learning Objectives
1. Make judgments about the size of the standard error of the estimate from a scatter plot
2. Compute the standard error of the estimate based on errors of prediction
3. Compute the standard error using Pearson's correlation
4. Estimate the standard error of the estimate based on a sample

Figure 1 shows two regression examples. You can see that in Graph A, the points are closer to the line than they are in Graph B. Therefore, the predictions in Graph A are more accurate than in Graph B.

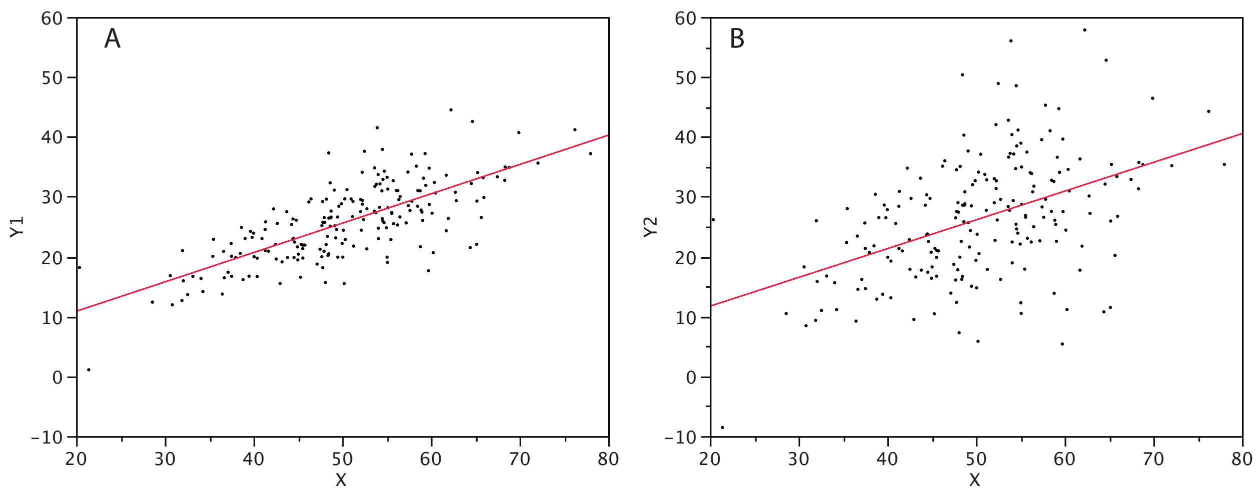


Figure 1. Regressions differing in accuracy of prediction.

The standard error of the estimate is a measure of the accuracy of predictions. Recall that the regression line is the line that minimizes the sum of squared deviations of prediction (also called the sum of squares error). The standard error of the estimate is closely related to this quantity and is defined below:

$$\sigma_{est} = \sqrt{\frac{\sum (Y - Y')^2}{N}}$$

where σ_{est} is the standard error of the estimate, Y is an actual score, Y' is a predicted score, and N is the number of pairs of scores. The numerator is the sum of squared differences between the actual scores and the predicted scores.

Note the similarity of the formula for σ_{est} to the formula for σ:

$$\sigma = \sqrt{\frac{\sum (Y - \mu)^2}{N}}$$

In fact, σ_{est} is the standard deviation of the errors of prediction (each Y - Y' is an error of prediction).

Assume the data in Table 1 are the data from a population of five X, Y pairs.

Table 1. Example data.

	X	Y	Y'	Y-Y'	$(Y-Y')^2$
	1.00	1.00	1.210	-0.210	0.044
	2.00	2.00	1.635	0.365	0.133
	3.00	1.30	2.060	-0.760	0.578
	4.00	3.75	2.485	1.265	1.600
	5.00	2.25	2.910	-0.660	0.436
Sum	15.00	10.30	10.30	0.000	2.791

The last column shows that the sum of the squared errors of prediction is 2.791. Therefore, the standard error of the estimate is

$$\sigma_{est} = \sqrt{\frac{2.791}{5}} = 0.747$$

There is a version of the formula for the standard error in terms of Pearson's correlation:

$$\sigma_{est} = \sqrt{\frac{(1-\rho^2)SSY}{N}}$$

where ϱ is the population value of Pearson's correlation and SSY is

$$SSY = \sum (Y - \mu_Y)^2$$

For the data in Table 1, m_y = 10.30, SSY = 4.597 and r = 0.6268. Therefore,

$$\sigma_{est} = \sqrt{\frac{(1-0.6268^2)(4.597)}{5}} = \sqrt{\frac{2.791}{5}} = 0.747$$

which is the same value computed previously.

Similar formulas are used when the standard error of the estimate is computed from a sample rather than a population. The only difference is that the denominator is N-2 rather than N. The reason N-2 is used rather than N-1 is that two parameters (the slope and the intercept) were estimated in order to estimate the sum of squares. Formulas for a sample comparable to the ones for a population are shown below:

$$s_{est} = \sqrt{\frac{\sum (Y - Y')^2}{N-2}}$$

$$s_{est} = \sqrt{\frac{2.791}{3}} = 0.964$$

$$s_{est} = \sqrt{\frac{(1-r^2)SSY}{N-2}}$$

Inferential Statistics for b and r

by David M. Lane

Prerequisites
• Chapter 9: Sampling Distribution of r
• Chapter 9: Confidence Interval for r

Learning Objectives
1. State the assumptions that inferential statistics in regression are based upon
2. Identify heteroscedasticity in a scatter plot
3. Compute the standard error of a slope
4. Test a slope for significance
5. Construct a confidence interval on a slope
6. Test a correlation for significance
7. Construct a confidence interval on a correlation

This section shows how to conduct significance tests and compute confidence intervals for the regression slope and Pearson's correlation. As you will see, if the regression slope is significantly different from zero, then the correlation coefficient is also significantly different from zero.

Assumptions

Although no assumptions were needed to determine the best-fitting straight line, assumptions are made in the calculation of inferential statistics. Naturally, these assumptions refer to the population, not the sample.

1. Linearity: The relationship between the two variables is linear.
2. Homoscedasticity: The variance around the regression line is the same for all values of X. A clear violation of this assumption is shown in Figure 1. Notice that the predictions for students with high high-school GPAs are very good, whereas the predictions for students with low high-school GPAs are not very good. In other words, the points for students with high high-school GPAs are close to the regression line, whereas the points for low high-school GPA students are not.

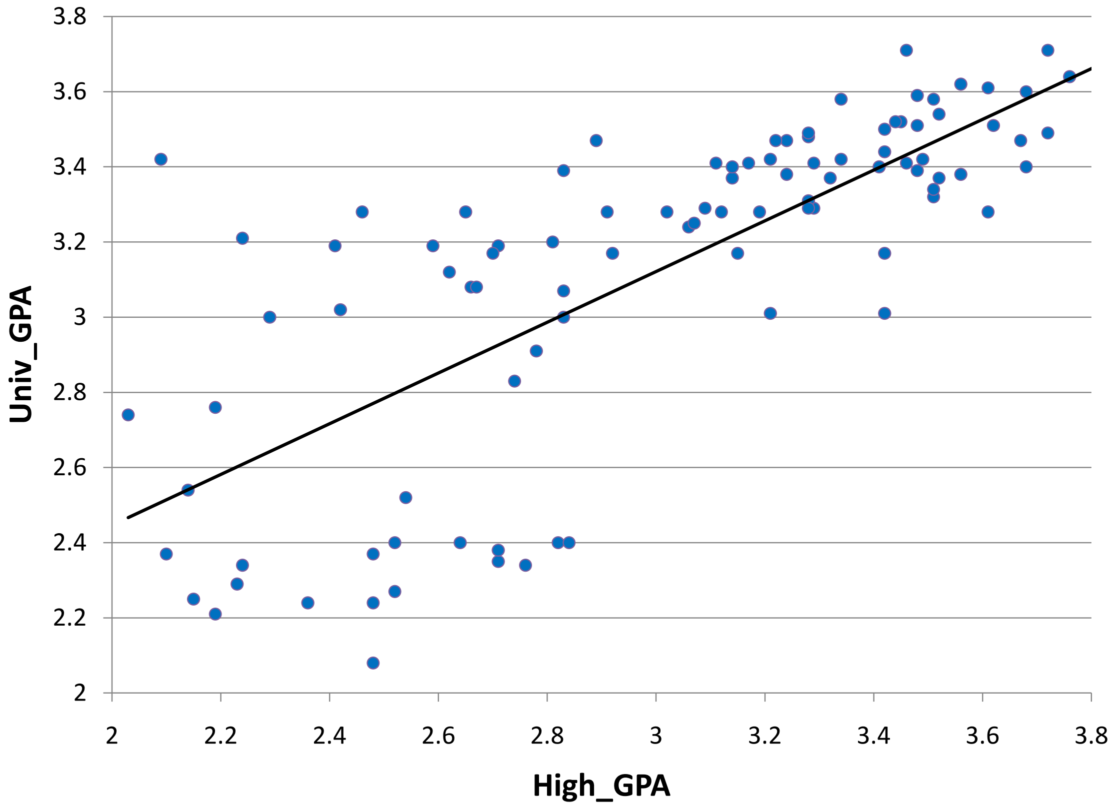


Figure 1. University GPA as a function of High School GPA.

3. The errors of prediction are distributed normally. This means that the distributions of deviations from the regression line are normally distributed. It does not mean that X or Y is normally distributed.

Significance Test for the Slope (b)

Recall the general formula for a t test:

$$t = \frac{statistics - hypothesized\ value}{estimated\ standard\ error\ of\ the\ statistic}$$

As applied here, the statistic is the sample value of the slope (b) and the hypothesized value is 0. The degrees of freedom for this test are:

```
df = N-2
```

where N is the number of pairs of scores.

The estimated standard error of b is computed using the following formula:

$$s_b = \frac{s_{est}}{\sqrt{SSX}}$$

where s_b is the estimated standard error of b, s_{est} is the standard error of the estimate, and SSX is the sum of squared deviations of X from the mean of X. SSX is calculated as

$$SSX = \sum (X - M_x)^2$$

where M_x is the mean of X. As shown previously, the standard error of the estimate can be calculated as

$$s_{est} = \sqrt{\frac{(1 - r^2)SSY}{N - 2}}$$

These formulas are illustrated with the data shown in Table 1. These data are reproduced from the introductory section. The column X has the values of the *predictor variable* and the column Y has the values of the *criterion variable*. The third column, x, contains the differences between the values of column X and the mean of X. The fourth column, x^2, is the square of the x column. The fifth column, y, contains the differences between the values of column Y and the mean of Y. The last column, y^2, is simply the square of the y column.

Table 1. Example data.

	X	Y	x	x^2	y	y^2
	1.00	1.00	-2.00	4	-1.06	1.1236
	2.00	2.00	-1.00	1	-0.06	0.0036
	3.00	1.30	0.00	0	-0.76	0.5776
	4.00	3.75	1.00	1	1.69	2.8561
	5.00	2.25	2.00	4	0.19	0.0361
Sum	15.00	10.30	0.00	10.00	0.00	4.5970

The computation of the standard error of the estimate (s_{est}) for these data is shown in the section on the standard error of the estimate. It is equal to 0.964.

$$s_{est} = 0.964$$

SSX is the sum of squared deviations from the mean of X. It is, therefore, equal to the sum of the x^2 column and is equal to 10.

$$SSX = 10.00$$

We now have all the information to compute the standard error of b:

$$s_b = \frac{0.964}{\sqrt{10}} = 0.305$$

As shown previously, the slope (b) is 0.425. Therefore,

$$t = \frac{0.425}{0.305} = 1.39$$

$$df = N-2 = 5-2 = 3.$$

The p value for a two-tailed t test is 0.26. Therefore, the slope is not significantly different from 0.

Confidence Interval for the Slope

The method for computing a confidence interval for the population slope is very similar to methods for computing other confidence intervals. For the 95% confidence interval, the formula is:

$$\text{lower limit: } b - (t_{.95})(s_b)$$
$$\text{upper limit: } b + (t_{.95})(s_b)$$

where $t_{.95}$ is the value of t to use for the 95% confidence interval.

The values of t to be used in a confidence interval can be looked up in a table of the t distribution. A small version of such a table is shown in Table 2. The first column, df, stands for degrees of freedom.

Table 2. Abbreviated t table.

df	0.95	0.99
2	4.303	9.925
3	3.182	5.841
4	2.776	4.604
5	2.571	4.032
8	2.306	3.355
10	2.228	3.169
20	2.086	2.845
50	2.009	2.678
100	1.984	2.626

You can also use the "inverse t distribution" calculator (external link; requires Java) to find the t values to use in a confidence interval.

Applying these formulas to the example data,

$$\text{lower limit: } 0.425 - (3.182)(0.305) = -0.55$$
$$\text{upper limit: } 0.425 + (3.182)(0.305) = 1.40$$

Significance Test for the Correlation

The formula for a significance test of Pearson's correlation is shown below:

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

where N is the number of pairs of scores. For the example data,

$$t = \frac{0.627\sqrt{5-2}}{\sqrt{1-0.627^2}} = 1.39$$

Notice that this is the same t value obtained in the t test of b. As in that test, the degrees of freedom is N-2 = 5-2 = 3.

Influential Observations

by David M. Lane

Prerequisites
• Chapter 14: Introduction to Linear Regression

Learning Objectives
1. Define “influence”
2. Describe what makes a point influential
3. Define “leverage”
4. Define “distance”

It is possible for a single observation to have a great influence on the results of a regression analysis. It is therefore important to be alert to the possibility of influential observations and to take them into consideration when interpreting the results.

Influence

The influence of an observation can be thought of in terms of how much the predicted scores for other observations would differ if the observation in question were not included. Cook's D is a good measure of the influence of an observation and is proportional to the sum of the squared differences between predictions made with all observations in the analysis and predictions made leaving out the observation in question. If the predictions are the same with or without the observation in question, then the observation has no influence on the regression model. If the predictions differ greatly when the observation is not included in the analysis, then the observation is influential.

A common rule of thumb is that an observation with a value of Cook's D over 1.0 has too much influence. As with all rules of thumb, this rule should be applied judiciously and not thoughtlessly.

An observation's influence is a function of two factors: (1) how much the observation's value on the predictor variable differs from the mean of the predictor variable and (2) the difference between the predicted score for the observation and its actual score. The former factor is called the observation's leverage. The latter factor is called the observation's *distance*.

Calculation of Cook's D (Optional)

The first step in calculating the value of Cook's D for an observation is to predict all the scores in the data once using a regression equation based on all the observations and once using all the observations except the observation in question. The second step is to compute the sum of the squared differences between these two sets of predictions. The final step is to divide this result by 2 times the MSE (see the section on partitioning the variance).

Leverage

The *leverage* of an observation is based on how much the observation's value on the predictor variable differs from the mean of the predictor variable. The greater an observation's leverage, the more potential it has to be an influential observation. For example, an observation with the mean on the predictor variable has no influence on the slope of the regression line regardless of its value on the criterion variable. On the other hand, an observation that is extreme on the predictor variable has, depending on its distance, the potential to affect the slope greatly.

Calculation of Leverage (h)

The first step is to standardize the predictor variable so that it has a mean of 0 and a standard deviation of 1. Then, the leverage (h) is computed by squaring the observation's value on the standardized predictor variable, adding 1, and dividing by the number of observations.

Distance

The distance of an observation is based on the error of prediction for the observation: The greater the error of prediction, the greater the distance. The most commonly used measure of distance is the *studentized residual*. The studentized residual for an observation is closely related to the error of prediction for that observation divided by the standard deviation of the errors of prediction. However, the predicted score is derived from a regression equation in which the observation in question is not counted. The details of the computation of a studentized residual are a bit complex and are beyond the scope of this work.

An observation with a large distance will not have that much influence if its leverage is low. It is the combination of an observation's leverage and distance that determines its influence.

Example

Table 1 shows the leverage, studentized residual, and influence for each of the five observations in a small dataset.

Table 1. Example Data.

ID	X	Y	h	R	D
A	1	2	0.39	-1.02	0.40
B	2	3	0.27	-0.56	0.06
C	3	5	0.21	0.89	0.11
D	4	6	0.20	1.22	0.19
E	8	7	0.73	-1.68	8.86
h is the leverage, R is the studentized residual, and D is Cook's measure of influence.					

Observation A has fairly high leverage, a relatively high residual, and moderately high influence.

Observation B has small leverage and a relatively small residual. It has very little influence.

Observation C has small leverage and a relatively high residual. The influence is relatively low.

Observation D has the lowest leverage and the second highest residual. Although its residual is much higher than Observation A, its influence is much less because of its low leverage.

Observation E has by far the largest leverage and the largest residual. This combination of high leverage and high residual makes this observation extremely influential.

Figure 1 shows the regression line for the whole dataset (blue) and the regression line if the observation in question is not included (red) for all observations. The observation in question is circled. Naturally, the regression line for the whole dataset is the same in all panels. The residual is calculated relative to the line for which the observation in question is not included in the analysis. This can be seen most clearly for Observation E which lies very close to the regression line computed when it is included but very far from the regression line when it is excluded from the calculation of the line.

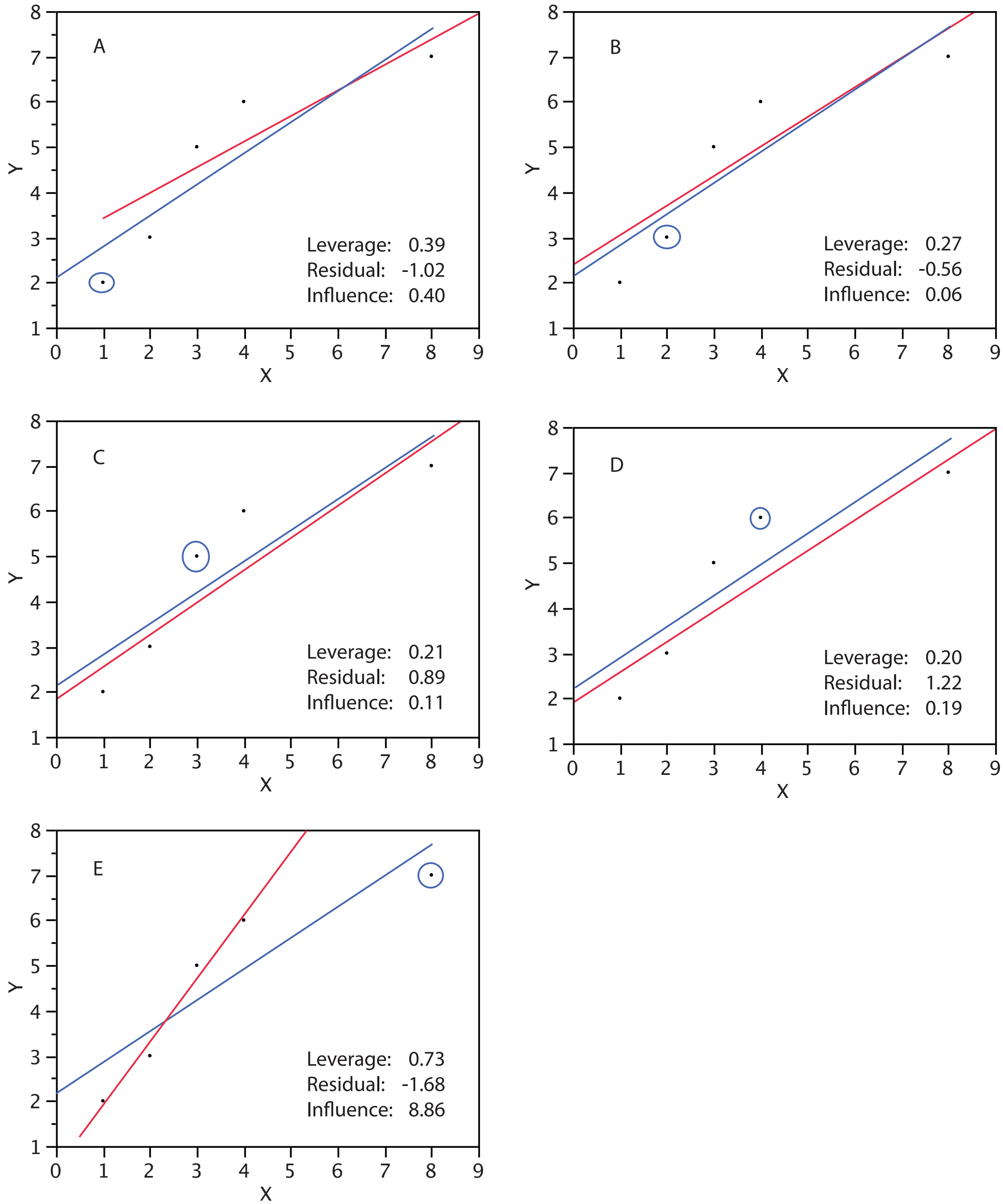


Figure 1. Illustration of leverage, residual, and influence.

The circled points are not included in the calculation of the red regression line. All points are included in the calculation of the blue regression line.

Regression Toward the Mean

by David M. Lane

Prerequisites
• Chapter 14: Regression Introduction

Learning Objectives
1. Explain what regression towards the mean is
2. State the conditions under which regression toward the mean occurs
3. Identify situations in which neglect of regression toward the mean leads to incorrect conclusions
4. Explain how regression toward the mean relates to a regression equation.

Regression toward the mean involves outcomes that are at least partly due to chance. We begin with an example of a task that is entirely chance: Imagine an experiment in which a group of 25 people each predicted the outcomes of flips of a fair coin. For each subject in the experiment, a coin is flipped 12 times and the subject predicts the outcome of each flip. Figure 1 shows the results of a simulation of this "experiment." Although most subjects were correct from 5 to 8 times out of 12, one simulated subject was correct 10 times. Clearly, this subject was very lucky and probably would not do as well if he or she performed the task a second time. In fact, the best prediction of the number of times this subject would be correct on the retest is 6 since the probability of being correct on a given trial is 0.5 and there are 12 trials.

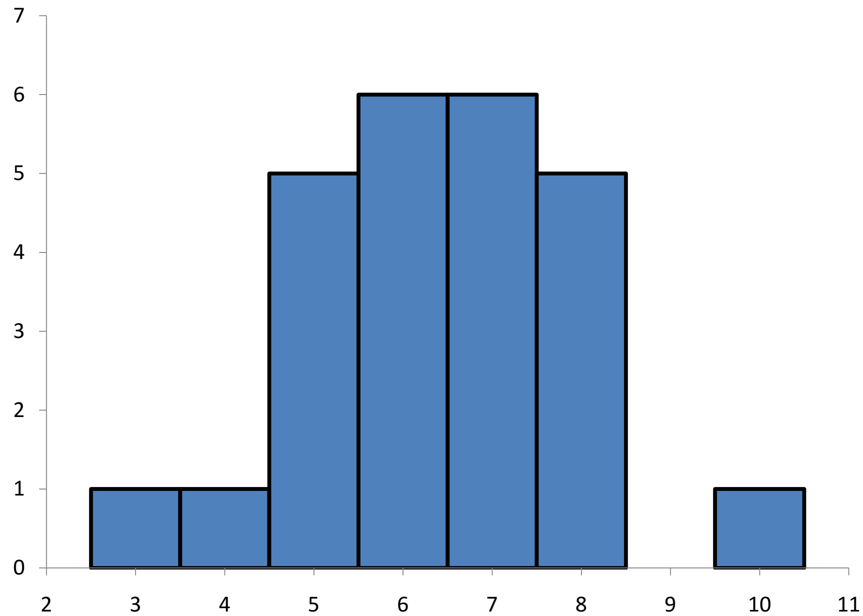


Figure 1. Histogram of results of a simulated experiment.

More technically, the best prediction for the subject's result on the retest is the mean of the binomial distribution with N = 12 and p = 0.50. This distribution is shown in Figure 2 and has a mean of 6.

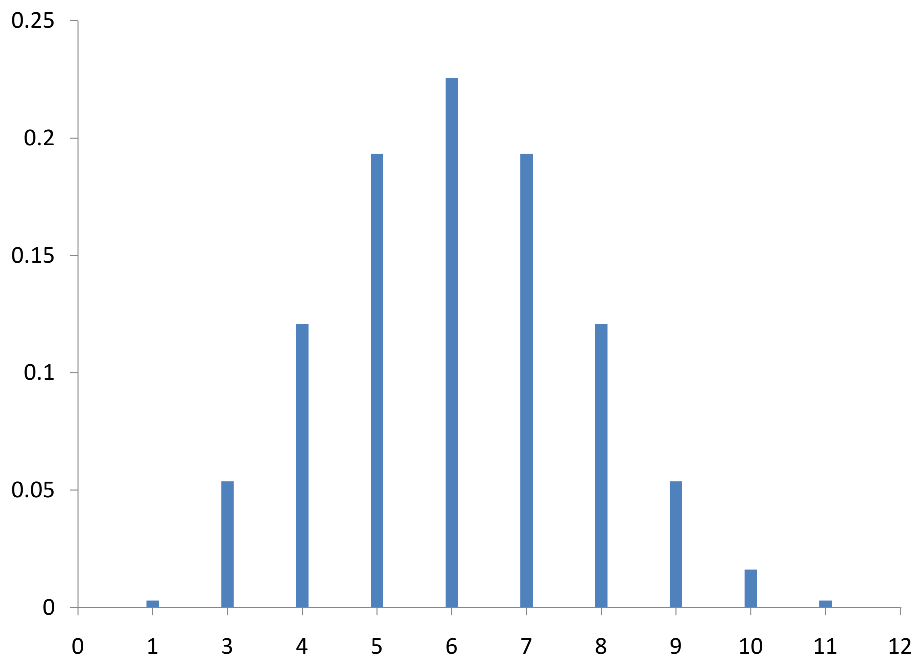


Figure 2. Binomial Distribution for N = 12 and p = .50.

The point here is that no matter how many coin flips a subject predicted correctly, the best prediction of their score on a retest is 6.

Now we consider a test we will call “Test A” that is partly chance and partly skill: Instead of predicting the outcomes of 12 coin flips, each subject predicts the outcomes of 6 coin flips and answers 6 true/false questions about world history. Assume that the mean score on the 6 history questions is 4. A subject's score on Test A has a large chance component but also depends on history knowledge. If a subject scored very high on this test (such as a score of 10/12), it is likely that they did well on both the history questions and the coin flips. For example, if they only got four of the history questions correct, they would have had to have gotten all six of the coin predictions correct, and this would have required exceptionally good luck. If given a second test (Test B) that also included coin predictions and history questions, their knowledge of history would be helpful and they would again be expected to score above the mean. However, since their high performance on the coin portion of Test A would not be predictive of their coin performance on Test B, they would not be expected to fare as well on Test B as on Test A. Therefore, the best prediction of their score on Test B would be somewhere between their score on Test A and the mean of Test B. This tendency of subjects with high values on a measure that includes chance and skill to score closer to the mean on a retest is called “*regression toward the mean*.”

The essence of the regression-toward-the-mean phenomenon is that people with high scores tend to be above average in skill and in luck, and that only the skill portion is relevant to future performance. Similarly, people with low scores tend to be below average in skill and luck and their bad luck is not relevant to future performance. This does not mean that all people who score high have above average luck. However, on average they do.

Almost every measure of behavior has a chance and a skill component to it. Take a student's grade on a final exam as an example. Certainly, the student's knowledge of the subject will be a major determinant of his or her grade. However, there are aspects of performance that are due to chance. The exam cannot cover everything in the course and therefore must represent a subset of the material. Maybe the student was lucky in that the one aspect of the course the student did not understand well was not well represented on the test. Or, maybe, the student was not sure which of two approaches to a problem would be better but, more or less by chance, chose the right one. Other chance elements come into play as well. Perhaps the student was awakened early in the morning by a random phone call, resulting in fatigue and lower performance. And, of course, guessing on multiple choice questions is another source of randomness in test scores.

There will be regression toward the mean in a test-retest situation whenever there is less than a perfect (r = 1) relationship between the test and the retest. This follows from the formula for a regression line with standardized variables shown below.

$$Z_{Y'} = (r)(Z_X)$$

From this equation it is clear that if the absolute value of r is less than 1, then the predicted value of Z_Y will be closer to 0, the mean for standardized scores, than is Z_X. Also, note that if the correlation between X and Y is 0, as it would be for a task that is all luck, the predicted standard score for Y is its mean, 0, regardless of the score on X.

Figure 3 shows a scatter plot with the regression line predicting the standardized Verbal SAT from the standardized Math SAT. Note that the slope of the line is equal to the correlation of 0.835 between these variables.

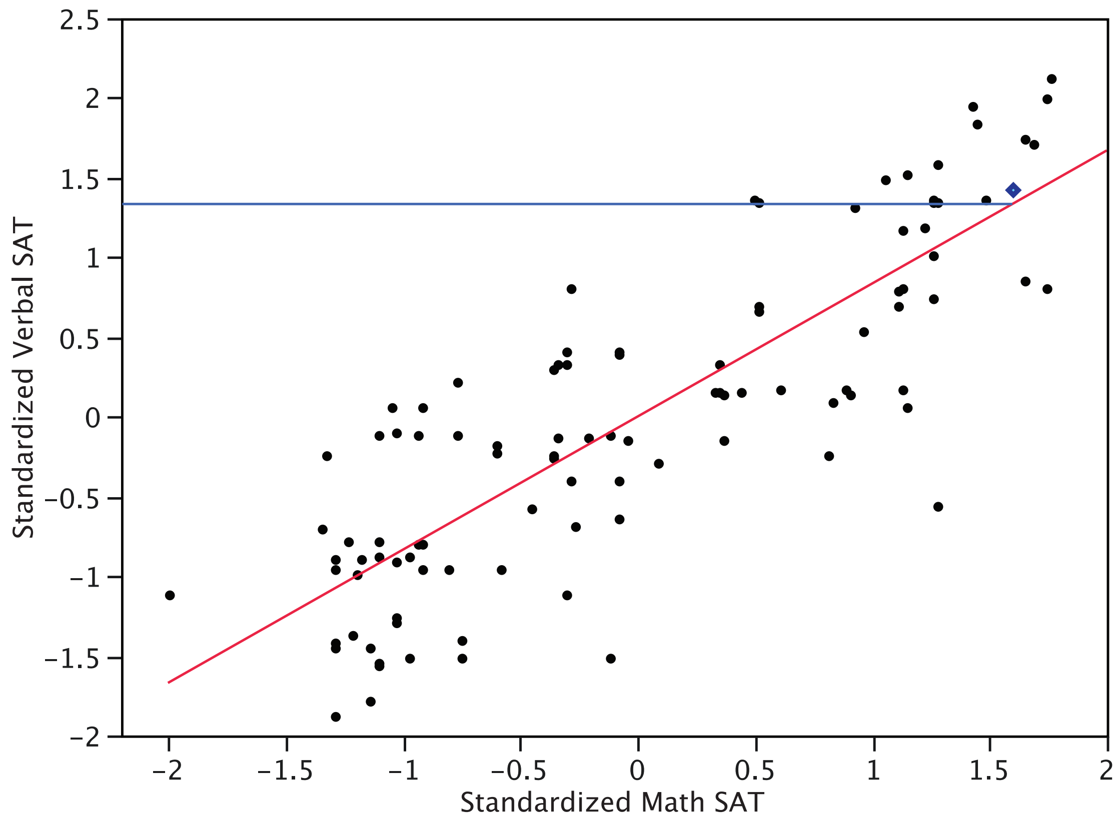


Figure 3. Prediction of Standardized Verbal SAT from Standardized Math SAT.

The point represented by a blue diamond has a value of 1.6 on the standardized Math SAT. This means that this student scored 1.6 standard deviations above the mean on Math SAT. The predicted score is (r)(1.6) = (0.835)(1.6) = 1.34. The horizontal line on the graph shows the value of the predicted score. The key point is that although this student scored 1.6 standard deviations above the mean on Math SAT, he or she is only predicted to score 1.34 standard deviations above the mean on Verbal SAT. Thus, the prediction is that the Verbal SAT score will be closer to the mean of 0 than is the Math SAT score. Similarly, a student scoring far below the mean on Math SAT will be predicted to score higher on Verbal SAT.

Regression toward the mean occurs in any situation in which observations are selected on the basis of performance on a task that has a random component. If you choose people on the basis of their performance on such a task, you will be choosing people partly on the basis of their skill and partly on the basis of their luck on the task. Since their luck cannot be expected to be maintained from trial to trial, the best prediction of a person's performance on a second trial will be somewhere between their performance on the first trial and the mean performance on the first trial. The degree to which the score is expected to "regress toward the mean" in this manner depends on the relative contributions of chance and skill to the task: the greater the role of chance, the more the regression toward the mean.

Errors Resulting From Failure to Understand Regression Toward the Mean

Failure to appreciate regression toward the mean is common and often leads to incorrect interpretations and conclusions. One of the best examples is provided by Nobel Laureate Daniel Kahneman in his autobiography (external link). Dr. Kahneman was attempting to teach flight instructors that praise is more effective than punishment. He was challenged by one of the instructors who relayed that in his experience praising a cadet for executing a clean maneuver is typically followed by a lesser performance, whereas screaming at a cadet for bad execution is typically followed by improved performance. This, of course, is exactly what would be expected based on regression toward the mean. A pilot's performance, although based on considerable skill, will vary randomly from maneuver to maneuver. When a pilot executes an extremely clean maneuver, it is likely that he or she had a bit of luck in their favor in addition to their considerable skill. After the praise but not because of it, the luck component will probably disappear and the performance will be lower. Similarly, a poor performance is likely to be partly due to bad luck. After the criticism but not because of it, the next performance will likely be better. To drive this point home, Kahneman had each instructor perform a task in which a coin was tossed at a target twice. He demonstrated that the performance of those who had done the best the first time deteriorated, whereas the performance of those who had done the worst improved.

Regression toward the mean is frequently present in sports performance. A good example is provided by Schall and Smith (2000), who analyzed many aspects of baseball statistics including the batting averages of players in 1998. They chose the 10 players with the highest batting averages (BAs) in 1998 and checked to see how well they did in 1999. According to what would be expected based on regression toward the mean, these players should, on average, have lower batting averages in 1999 than they did in 1998. As can be seen in Table 1, 7/10 of the players had lower batting averages in 1999 than they did in 1998. Moreover, those who had higher averages in 1999 were only slightly higher, whereas those who were lower were much lower. The average decrease from 1998 to 1999 was 33 points. Even so, most of these players had excellent batting averages in 1999 indicating that skill was an important component of their 1998 averages.

Table 1. How the Ten Players with the Highest BAs in 1998 did in 1999.

1998	1999	Difference
363	379	16
354	298	-56
339	342	3
337	281	-56
336	249	-87
331	298	-33
328	297	-31
328	303	-25
327	257	-70
327	332	5

Figure 4 shows the batting averages of the two years. The decline from 1998 to 1999 is clear. Note that although the mean decreased from 1998, some players increased their batting averages. This illustrates that regression toward the mean does not occur for every individual. Although the predicted scores for every individual will be lower, some of the predictions will be wrong.

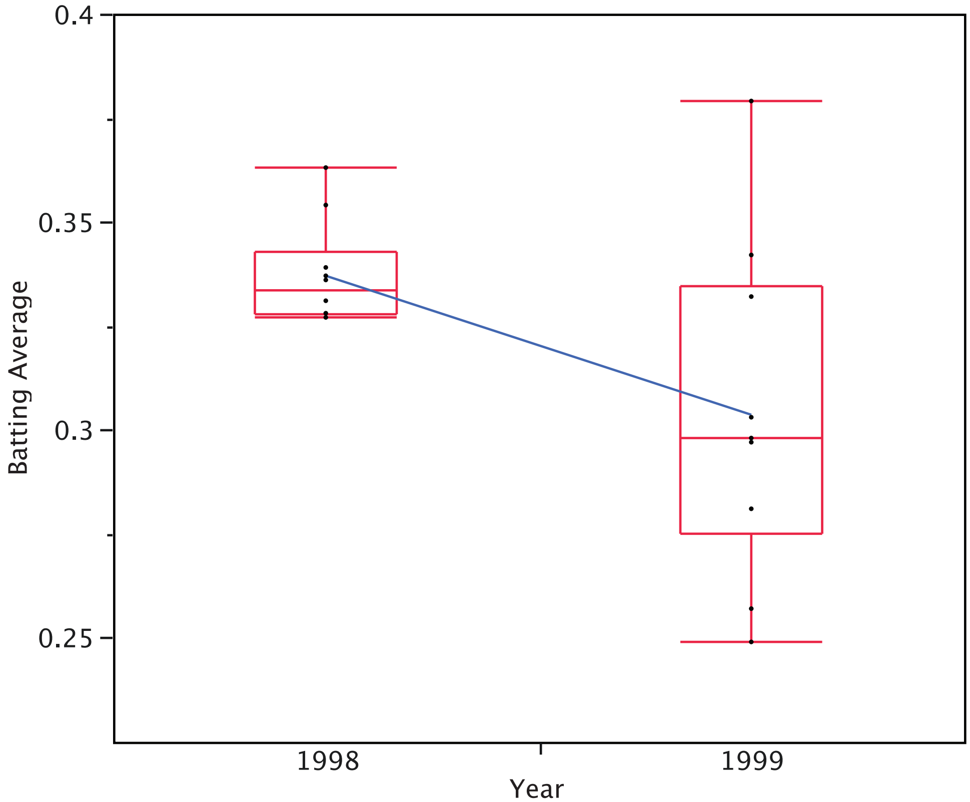


Figure 4. Quantile plots of the batting averages. The line connects the means of the plots.

Regression toward the mean plays a role in the so-called “Sophomore Slump,” a good example of which is that a player who wins “rookie of the year” typically does less well in his second season. A related phenomenon is called the Sports Illustrated Cover Jinx.

An experiment without a control group can *confound* regression effects with real effects. For example, consider a hypothetical experiment to evaluate a reading-improvement program. All first graders in a school district were given a reading achievement test and the 50 lowest-scoring readers were enrolled in the program. The students were retested following the program and the mean improvement was large. Does this necessarily mean the program was effective? No, it could be that the initial poor performance of the students was due, in part, to bad luck. Their luck would be expected to improve in the retest, which would increase their scores with or without the treatment program.

For a real example, consider an experiment that sought to determine whether the drug propranolol would increase the SAT scores of students thought to have test anxiety (external link). Propranolol was given to 25 high-school students chosen because IQ tests and other academic performance indicated that they had not done as well as expected on the SAT. On a retest taken after receiving propranolol, students improved their SAT scores an average of 120 points. This was a *significantly* greater increase than the 38 points expected simply on the basis of having taken the test before. The problem with the study is that the method of selecting students likely resulted in a disproportionate number of students who had bad luck when they first took the SAT. Consequently, these students would likely have increased their scores on a retest with or without the propranolol. This is not to say that propranolol had no effect. However, since possible propranolol effects and regression effects were confounded, no firm conclusions should be drawn.

Randomly assigning students to either the propranolol group or a control group would have improved the experimental design. Since the regression effects would then not have been systematically different for the two groups, a significant difference would have provided good evidence for a propranolol effect.

Introduction to Multiple Regression

by David M. Lane

Prerequisites
- Chapter 14: Simple Linear Regression
- Chapter 14: Partitioning Sums of Squares
- Chapter 14: Standard Error of the Estimate
- Chapter 14: Inferential Statistics for b and r

Learning Objectives
1. State the regression equation
2. Define “regression coefficient”
3. Define “beta weight”
4. Explain what R is and how it is related to r
5. Explain why a regression weight is called a “partial slope”
6. Explain why the sum of squares explained in a multiple regression model is usually less than the sum of the sums of squares in simple regression
7. Define R^2 in terms of proportion explained
8. Test R^2 for significance
9. Test the difference between a complete and reduced model for significance
10. State the assumptions of multiple regression and specify which aspects of the analysis require assumptions

In simple linear regression, a criterion variable is predicted from one predictor variable. In multiple regression, the criterion is predicted by two or more variables. For example, in the SAT case study, you might want to predict a student's university grade point average on the basis of their High-School GPA (HSGPA) and their total SAT score (verbal + math). The basic idea is to find a linear combination of HSGPA and SAT that best predicts University GPA (UGPA). That is, the problem is to find the values of b_1 and b_2 in the equation shown below that gives the best predictions of UGPA. As in the case of simple linear regression, we define the best predictions as the predictions that minimize the squared errors of prediction.

$$UGPA' = b_1HSGPA + b_2SAT + A$$

where UGPA' is the predicted value of University GPA and A is a constant. For these data, the best prediction equation is shown below:

```
UGPA' = 0.541 x HSGPA + 0.008 x SAT + 0.540
```

In other words, to compute the prediction of a student's University GPA, you add up (a) their High-School GPA multiplied by 0.541, (b) their SAT multiplied by 0.008, and (c) 0.540. Table 1 shows the data and predictions for the first five students in the dataset.

Table 1. Data and Predictions.

HSGPA	SAT	UGPA'
3.45	1232	3.38
2.78	1070	2.89
2.52	1086	2.76
3.67	1287	3.55
3.24	1130	3.19

The values of b (b1 and b2) are sometimes called “regression coefficients” and sometimes called “*regression weights*.” These two terms are synonymous.

The *multiple correlation* (R) is equal to the correlation between the predicted scores and the actual scores. In this example, it is the correlation between UGPA' and UGPA, which turns out to be 0.79. That is, R = 0.79. Note that R will never be negative since if there are negative correlations between the predictor variables and the criterion, the regression weights will be negative so that the correlation between the predicted and actual scores will be positive.

Interpretation of Regression Coefficients

A regression coefficient in multiple regression is the slope of the linear relationship between the criterion variable and the part of a predictor variable that is independent of all other predictor variables. In this example, the regression coefficient for HSGPA can be computed by first predicting HSGPA from SAT and saving the errors of prediction (the differences between HSGPA and HSGPA'). These errors of prediction are called “residuals” since they are what is left over in HSGPA after the predictions from SAT are subtracted, and they represent the part of HSGPA that is independent of SAT. These residuals are referred to as HSGPA.SAT, which means they are the residuals in HSGPA after having been predicted by SAT. The correlation between HSGPA.SAT and SAT is necessarily 0.

The final step in computing the regression coefficient is to find the slope of the relationship between these residuals and UGPA. This slope is the regression coefficient for HSGPA. The following equation is used to predict HSGPA from SAT:

```
HSGPA' = -1.314 + 0.0036 x SAT
```

The residuals are then computed as:

```
HSGPA - HSGPA'
```

The linear regression equation for the prediction of UGPA by the residuals is

```
UGPA' = 0.541 x HSGPA.SAT + 3.173
```

Notice that the slope (0.541) is the same value given previously for b_1 in the multiple regression equation.

This means that the regression coefficient for HSGPA is the slope of the relationship between the criterion variable and the part of HSPGA that is *independent* of (uncorrelated with) the other predictor variables. It represents the change in the criterion variable associated with a change of one in the predictor variable when all other predictor variables are held constant. Since the regression coefficient for HSGPA is 0.54, this means that, holding SAT constant, a change of one in HSGPA is associated with a change of 0.54 in UGPA. If two students had the same SAT and differed in HSGPA by 2, then you would predict they would differ in UGPA by (2)(0.54) = 1.08. Similarly, if they differed by 0.5, then you would predict they would differ by (0.50)(0.54) = 0.27.

The slope of the relationship between the part of a predictor variable independent of other predictor variables and the criterion is its partial slope. Thus the regression coefficient of 0.541 for HSGPA and the regression coefficient of 0.008 for SAT are partial slopes. Each partial slope represents the relationship between the predictor variable and the criterion holding constant all of the other predictor variables.

It is difficult to compare the coefficients for different variables directly because they are measured on different scales. A difference of 1 in HSGPA is a fairly large difference, whereas a difference of 1 on the SAT is negligible. Therefore, it can be advantageous to transform the variables so that they are on the same scale. The most straightforward approach is to standardize the variables so that they all have a standard deviation of 1. A regression weight for standardized variables is called a "beta weight" and is designated by the Greek letter β. For these data, the beta weights are 0.625 and 0.198. These values represent the change in the criterion (in standard deviations) associated with a change of one standard deviation on a predictor [holding constant the value(s) on the other predictor(s)]. Clearly, a change of one standard deviation on HSGPA is associated with a larger difference than a change of one standard deviation of SAT. In practical terms, this means that if you know a student's HSGPA, knowing the student's SAT does not aid the prediction of UGPA much. However, if you do not know the student's HSGPA, his or her SAT can aid in the prediction since the β weight in the simple regression predicting UGPA from SAT is 0.68. For comparison purposes, the β weight in the simple regression predicting UGPA from HSGPA is 0.78. As is typically the case, the partial slopes are smaller than the slopes in simple regression.

Partitioning the Sums of Squares

Just as in the case of simple linear regression, the sum of squares for the criterion (UGPA in this example) can be partitioned into the sum of squares predicted and the sum of squares error. That is,

$$SSY = SSY' + SSE$$

which for these data:

$$20.798 = 12.961 + 7.837$$

The sum of squares predicted is also referred to as the "sum of squares explained." Again, as in the case of simple regression,

$$\text{Proportion Explained} = SSY'/SSY$$

In simple regression, the proportion of variance explained is equal to r^2; in multiple regression, the proportion of variance explained is equal to R^2.

In multiple regression, it is often informative to partition the sums of squares explained among the predictor variables. For example, the sum of squares explained for these data is 12.96. How is this value divided between HSGPA and SAT? One approach that, as will be seen, does not work is to predict UGPA in separate simple regressions for HSGPA and SAT. As can be seen in Table 2, the sum of squares in these separate simple regressions is 12.64 for HSGPA and 9.75 for SAT. If we add these two sums of squares we get 22.39, a value much larger than the sum of squares explained of 12.96 in the multiple regression analysis. The explanation is that HSGPA and SAT are highly correlated ($r = .78$) and therefore much of the variance in UGPA is confounded between HSGPA or SAT. That is, it could be explained by either HSGPA or SAT and is counted twice if the sums of squares for HSGPA and SAT are simply added.

Table 2. Sums of Squares for Various Predictors

Predictors	Sum of Squares
HSGPA	12.64
SAT	9.75
HSGPA and SAT	12.96

Table 3 shows the partitioning of the sums of squares into the sum of squares uniquely explained by each predictor variable, the sum of squares confounded between the two predictor variables, and the sum of squares error. It is clear from this table that most of the sum of squares explained is confounded between HSGPA and SAT. Note that the sum of squares uniquely explained by a predictor variable is analogous to the partial slope of the variable in that both involve the relationship between the variable and the criterion with the other variable(s) controlled.

Table 3. Partitioning the Sum of Squares

Source	Sum of Squares	Proportion
HSGPA (unique)	3.21	0.15
SAT (unique)	0.32	0.02
HSGPA and SAT (Confounded)	9.43	0.45
Error	7.84	0.38
Total	20.80	1.00

The sum of squares uniquely attributable to a variable is computed by comparing two regression models: the complete model and a reduced model. The complete model is the multiple regression with all the predictor variables included (HSGPA and SAT in this example). A reduced model is a model that leaves out one of the predictor variables. The sum of squares uniquely attributable to a variable is the sum of squares for the complete model minus the sum of squares for the reduced model in which the variable of interest is omitted. As shown in Table 2, the sum of squares for the complete model (HSGPA and SAT) is 12.96. The sum of squares for the reduced model in which HSGPA is omitted is simply the sum of squares explained using SAT as the predictor variable and is 9.75. Therefore, the sum of squares uniquely attributable to HSGPA is 12.96 - 9.75 = 3.21. Similarly, the sum of squares uniquely attributable to SAT is 12.96 - 12.64 = 0.32. The confounded sum of squares in this example is computed by subtracting the sum of squares uniquely attributable to the predictor variables from the sum of squares for the complete model: 12.96 - 3.21 - 0.32 = 9.43. The computation of the confounded sums of squares in analyses with more than two predictors is more complex and beyond the scope of this text.

Since the variance is simply the sum of squares divided by the degrees of freedom, it is possible to refer to the proportion of variance explained in the same way as the proportion of the sum of squares explained. It is slightly more common to refer to the proportion of variance explained than the proportion of the sum of squares explained and, therefore, that terminology will be adopted frequently here.

When variables are highly correlated, the variance explained uniquely by the individual variables can be small even though the variance explained by the variables taken together is large. For example, although the proportions of variance explained uniquely by HSGPA and SAT are only 0.15 and 0.02 respectively, together these two variables explain 0.62 of the variance. Therefore, you could easily underestimate the importance of variables if only the variance explained uniquely by each variable is considered. Consequently, it is often useful to consider a set of related variables. For example, assume you were interested in predicting job performance from a large number of variables some of which reflect cognitive ability. It is likely that these measures of cognitive ability would be highly correlated among themselves and therefore no one of them would explain much of the variance independent of the other variables. However, you could avoid this problem by determining the proportion of variance explained by all of the cognitive ability variables considered together as a set. The variance explained by the set would include all the variance explained uniquely by the variables in the set as well as all the variance confounded among variables in the set. It would not include variance confounded with variables outside the set. In short, you would be computing the variance explained by the set of variables that is independent of the variables not in the set.

Inferential Statistics

We begin by presenting the formula for testing the significance of the contribution of a set of variables. We will then show how special cases of this formula can be used to test the significance of R^2 as well as to test the significance of the unique contribution of individual variables.

The first step is to compute two regression analyses: (1) an analysis in which all the predictor variables are included and (2) an analysis in which the variables in the set of variables being tested are **excluded**. The former regression model is called the "complete model" and the latter is called the "reduced model." The basic idea is that if the reduced model explains much less than the complete model, then the set of variables excluded from the reduced model is important.

The formula for testing the contribution of a group of variables is:

$$F = \frac{\frac{SSQ_c - SSQ_R}{P_c - P_R}}{\frac{SSQ_T - SSQ_c}{N - P_c - 1}} = \frac{MS_{explained}}{MS_{error}}$$

where:

SSQ_C is the sum of squares for the complete model,

SSQ_R is the sum of squares for the reduced model,

p_C is the number of predictors in the complete model,

p_R is the number of predictors in the reduced model,

SSQ_T is the sum of squares total (the sum of squared deviations of the criterion variable from its mean), and

N is the total number of observations

The degrees of freedom for the numerator is p_c - p_r and the degrees of freedom for the denominator is N - p_C -1. If the F is significant, then it can be concluded that the variables excluded in the reduced set contribute to the prediction of the criterion variable independently of the other variables.

This formula can be used to test the significance of R^2 by defining the reduced model as having no predictor variables. In this application, SSQ_R and p_R = 0. The formula is then simplified as follows:

$$F_{(pc, N-pc-1)} = \frac{\dfrac{SSQ_C}{p_C}}{\dfrac{SSQ_T - SSQ_C}{N - p_C - 1}} = \frac{MS_{\text{explained}}}{MS_{error}}$$

which for this example becomes:

$$F = \frac{\dfrac{12.96}{2}}{\dfrac{20.80 - 12.96}{105 - 2 - 1}} = \frac{6.48}{0.08} = 84.35.$$

The degrees of freedom are 2 and 102. The F distribution calculator shows that $p < 0.001$.

The reduced model used to test the variance explained uniquely by a single predictor consists of all the variables except the predictor variable in question. For example, the reduced model for a test of the unique contribution of HSGPA contains only the variable SAT. Therefore, the sum of squares for the reduced model is the sum of squares when UGPA is predicted by SAT. This sum of squares is 9.75. The calculations for F are shown below:

$$F_{(1,102)} = \frac{\dfrac{12.96 - 9.75}{2 - 1}}{\dfrac{20.80 - 12.96}{105 - 2 - 1}} = \frac{3.212}{0.077} = 41.80.$$

The degrees of freedom are 1 and 102. The F distribution calculator shows that $p < 0.001$.

Similarly, the reduced model in the test for the unique contribution of SAT consists of HSGPA.

$$F = \frac{\dfrac{12.96 - 12.64}{2 - 1}}{\dfrac{20.80 - 12.96}{105 - 2 - 1}} = \frac{0.322}{0.077} = 4.19.$$

The degrees of freedom are 1 and 102. The F distribution calculator shows that $p = 0.0432$.

The significance test of the variance explained uniquely by a variable is identical to a significance test of the regression coefficient for that variable. A regression coefficient and the variance explained uniquely by a variable both reflect the relationship between a variable and the criterion independent of the other variables. If the variance explained uniquely by a variable is not zero, then the regression coefficient cannot be zero. Clearly, a variable with a regression coefficient of zero would explain no variance.

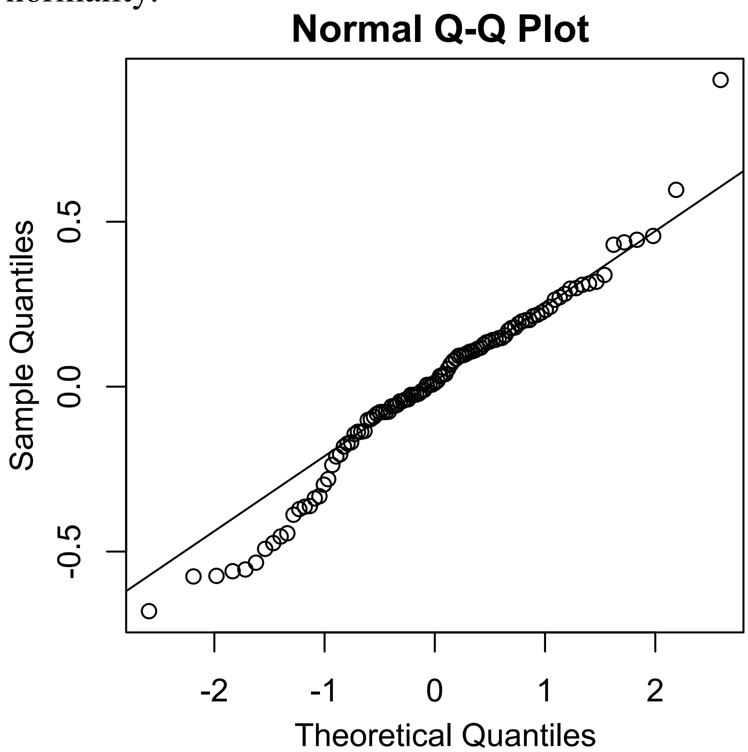
Other inferential statistics associated with multiple regression that are beyond the scope of this text. Two of particular importance are (1) confidence intervals on regression slopes and (2) confidence intervals on predictions for specific observations. These inferential statistics can be computed by standard statistical analysis packages such as R, SPSS, STATA, SAS, and JMP.

Assumptions

No assumptions are necessary for computing the regression coefficients or for partitioning the sums of squares. However, there are several assumptions made when interpreting inferential statistics. Moderate violations of Assumptions 1-3 do not pose a serious problem for testing the significance of predictor variables. However, even small violations of these assumptions pose problems for confidence intervals on predictions for specific observations.

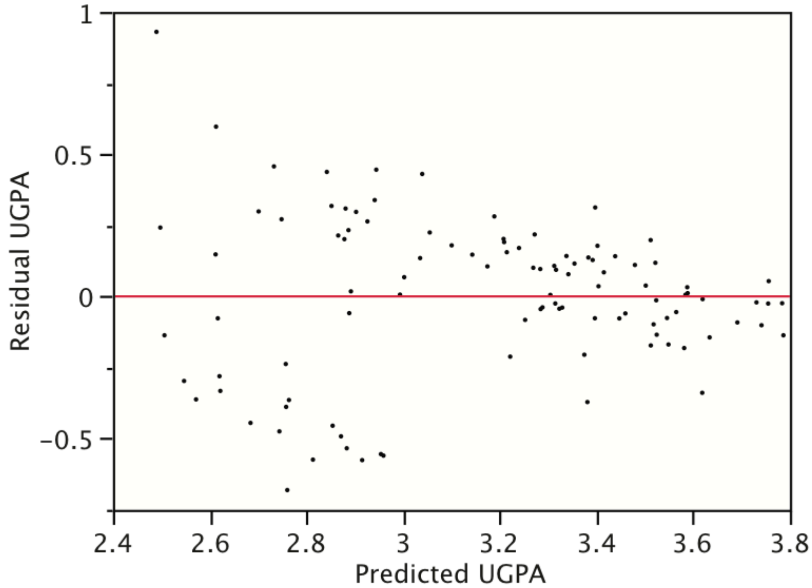
1. Residuals are normally distributed:
   As in the case of simple linear regression, the residuals are the errors of prediction. Specifically, they are the differences between the actual scores on the criterion and the predicted scores. A Q-Q plot for the residuals for the example data is shown below. This plot reveals that the actual data values at the lower end of the distribution do not increase as much as would be expected for a normal distribution. It also reveals that the highest value in the data is higher than would be expected for the highest value in a sample of this size from a normal distribution. Nonetheless, the distribution does not deviate greatly from

normality.



2. Homoscedasticity:
   It is assumed that the variance of the errors of prediction are the same for all predicted values. As can be seen below, this assumption is violated in the example data because the errors of prediction are much larger for observations with low-to-medium predicted scores than for observations with high predicted scores. Clearly, a confidence interval on a low predicted UGPA would

underestimate the uncertainty.



3. Linearity:
   It is assumed that the relationship between each predictor variable and the criterion variable is linear. If this assumption is not met, then the predictions may systematically overestimate the actual values for one range of values on a predictor variable and underestimate them for another.

Statistical Literacy

by David M. Lane

Prerequisites

- Chapter 14: Regression Toward the Mean

In a discussion about the Dallas Cowboy football team, there was a comment that the quarterback threw far more interceptions in the first two games than is typical (there were two interceptions per game). The author correctly pointed out that, because of regression toward the mean, performance in the future is expected to improve. However, the author defined regression toward the mean as, "In nerd land, that basically means that things tend to even out over the long run."

What do you think?

Comment on that definition.

> That definition is sort of correct, but it could be stated more precisely. Things don't always tend to even out in the long run. If a great player has an average game, then things wouldn't even out (to the average of all players) but would regress toward that player's high mean performance.

References

Schall, T., & Smith, G. (2000) Do Baseball Players Regress Toward the Mean? The *American Statistician*, 54, 231-235.