CK-12

flexbook
next generation textbooks

MAHS-DV Algebra 1-2 Q2

MAHS-DV Algebra 1-2 Q2

Adrienne Wooten

AUTHOR
Adrienne Wooten

To access a customizable version of this book, as well as other interactive content, visit www.ck12.org

CK-12 Foundation is a non-profit organization with a mission to reduce the cost of textbook materials for the K-12 market both in the U.S. and worldwide. Using an open-content, web-based collaborative model termed the **FlexBook®**, CK-12 intends to pioneer the generation and distribution of high-quality educational content that will serve both as core text as well as provide an adaptive environment for learning, powered through the **FlexBook Platform®**.

Printed: October 7, 2013

Contents

CHAPTER 1 Modeling Linear Data

Chapter Outline

1.1 Line Graphs

Here you'll learn the difference between continuous data and discrete data as it applies to a line graph. You'll also learn how to represent data that has a linear pattern on a graph and how to solve problems with line graphs.

You're scanning some photographs for a customer. The customer is charged a $25 set-up fee for the scanner and then 0.55 per scan. How much would the customer get charged for 8 scans? What about for 15 scans?

Watch This

First watch this video to learn about line graphs.

MEDIA
Click image to the left for more content.

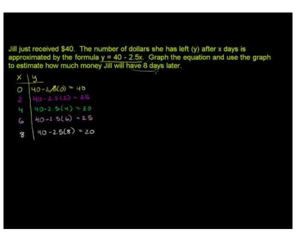
CK-12 Foundation: Chapter7LineGraphsA

Then watch this video to see some examples.

MEDIA
Click image to the left for more content.

CK-12 Foundation: Chapter7LineGraphsB

Watch this video for more help.



MEDIA
Click image to the left for more content.

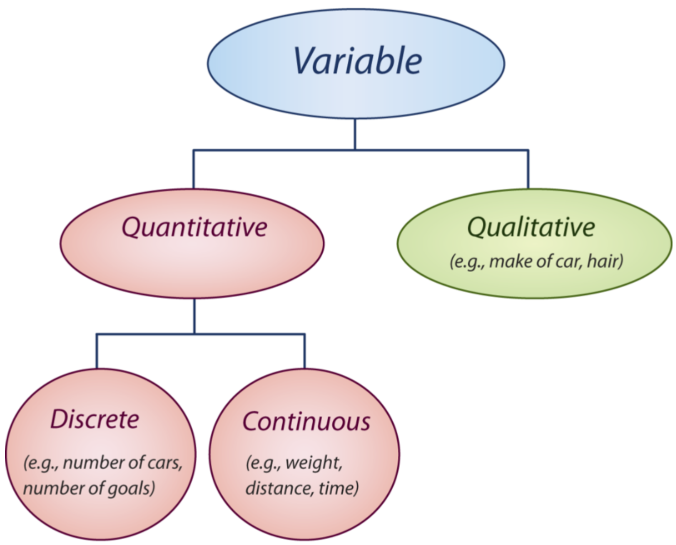
Khan Academy Application problem with graph

Guidance

Before you continue to explore the concept of representing data graphically, it is very important to understand the meaning of some basic terms that will often be used in this concept. The first such definition is that of a **variable**. In statistics, a variable is simply a characteristic that is being studied. This characteristic assumes different values for different elements, or members, of the population, whether it is the entire population or a sample. The value of the variable is referred to as an observation, or a measurement. A collection of these observations of the variable is a **data set**.

Variables can be quantitative or qualitative. A **quantitative variable** is one that can be measured numerically. Some examples of a quantitative variable are wages, prices, weights, numbers of vehicles, and numbers of goals. All of these examples can be expressed numerically. A quantitative variable can be classified as discrete or continuous. A **discrete variable** is one whose values are all countable and does not include any values between 2 consecutive values of a data set. An example of a discrete variable is the number of goals scored by a team during a hockey game. A **continuous variable** is one that can assume any countable value, as well as all the values between 2 consecutive numbers of a data set. An example of a continuous variable is the number of gallons of gasoline used during a trip to the beach.

A **qualitative variable** is one that cannot be measured numerically but can be placed in a category. Some examples of a qualitative variable are months of the year, hair color, color of cars, a person's status, and favorite vacation spots. The following flow chart should help you to better understand the above terms.



Variables can also be classified as dependent or independent. When there is a linear relationship between 2 variables, the values of one variable depend upon the values of the other variable. In a linear relation, the values of y depend upon the values of x. Therefore, the **dependent variable** is represented by the values that are plotted on the y-axis, and the **independent variable** is represented by the values that are plotted on the x-axis.

Example A

Select the best descriptions for the following variables and indicate your selections by marking an 'x' in the appropriate boxes.

TABLE 1.1:

Variable	Quantitative	Qualitative	Discrete	Continuous
Number of members in a family				
A person's marital status				
Length of a person's arm				

TABLE 1.1: (continued)

Variable	Quantitative	Qualitative	Discrete	Continuous
Color of cars				
Number of errors on a math test				

The variables can be described as follows:

TABLE 1.2:

Variable	Quantitative	Qualitative	Discrete	Continuous
Number of members in a family	x		x	
A person's marital status		x		
Length of a person's arm	x			x
Color of cars		x		
Number of errors on a math test	x		x	

Example B

Sally works at the local ballpark stadium selling lemonade. She is paid $15.00 each time she works, plus $0.75 for each glass of lemonade she sells. Create a table of values to represent Sally's earnings if she sells 8 glasses of lemonade. Use this table of values to represent her earnings on a graph.

The first step is to write an equation to represent her earnings and then to use this equation to create a table of values.

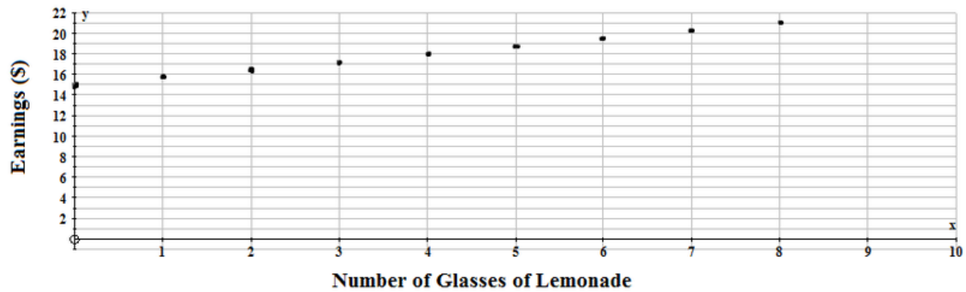
$y = 0.75x + 15$, where y represents her earnings and x represents the number of glasses of lemonade she sells.

TABLE 1.3:

Number of Glasses of Lemonade	Earnings
0	$15.00
1	$15.75
2	$16.50
3	$17.25
4	$18.00
5	$18.75
6	$19.50
7	$20.25
8	$21.00

The dependent variable is the money earned, and the independent variable is the number of glasses of lemonade sold. Therefore, money is on the y-axis, and the number of glasses of lemonade is on the x-axis.

From the table of values, Sally will earn $21.00 if she sells 8 glasses of lemonade.



Now that the points have been plotted, the decision has to be made as to whether or not to join them. Between every 2 points plotted on the graph are an infinite number of values. If these values are meaningful to the problem, then the plotted points can be joined. This type of data is called **continuous data**. If the values between the 2 plotted points are not meaningful to the problem, then the points should not be joined. This type of data is called **discrete data**. Since glasses of lemonade are represented by whole numbers, and since fractions or decimals are not appropriate values, the points between 2 consecutive values are not meaningful in this problem. Therefore, the points should not be joined. The data is discrete.

Guided Practice

The local arena is trying to attract as many participants as possible to attend the community's "Skate for Scoliosis" event. Participants pay a fee of $10.00 for registering, and, in addition, the arena will donate $3.00 for each hour a participant skates, up to a maximum of 6 hours. Create a table of values and draw a graph to represent a participant who skates for the entire 6 hours. How much money can a participant raise for the community if he/she skates for the maximum length of time?

Answer:

The equation for this scenario is $y = 3x + 10$, where y represents the money made by the participant, and x represents the number of hours the participant skates.

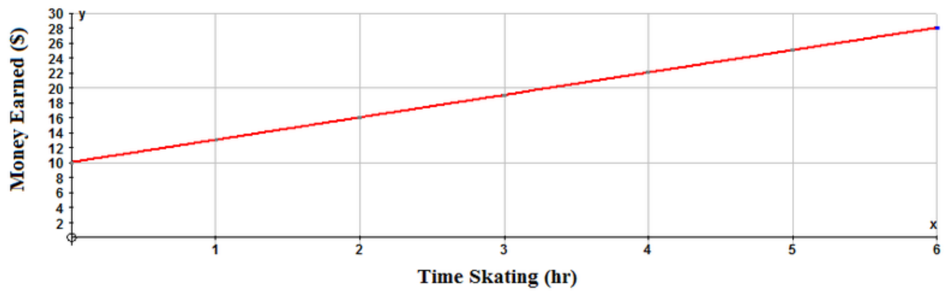
TABLE 1.4:

Numbers of Hours Skating	Money Earned
0	$10.00
1	$13.00

TABLE 1.4: (continued)

Numbers of Hours Skating	Money Earned
2	$16.00
3	$19.00
4	$22.00
5	$25.00
6	$28.00

The dependent variable is the money made, and the independent variable is the number of hours the participant skated. Therefore, money is on the y-axis, and time is on the x-axis as shown below:



A participant who skates for the entire 6 hours can make $28.00 for the "Skate for Scoliosis" event. The points are joined, because the fractions and decimals between 2 consecutive points are meaningful for this problem. A participant could skate for 30 minutes, and the arena would pay that skater $1.50 for the time skating. The data is continuous.

Interactive Practice

Practice

1. What term is used to describe a data set in which all points between 2 consecutive points are meaningful?
    a. discrete data
    b. continuous data
    c. random data
    d. fractional data
2. What type of variable is represented by the number of pets owned by families?
    a. qualitative
    b. quantitative
    c. independent
    d. continuous
3. What type of data, when plotted on a graph, does not have the points joined?
    a. discrete data
    b. continuous data
    c. random data
    d. independent data
4. Select the best descriptions for the following variables and indicate your selections by marking an 'x' in the appropriate boxes.

TABLE 1.5:

Variable	Quantitative	Qualitative	Discrete	Continuous
Men's favorite TV shows				
Salaries of baseball players				
Number of children in a family				
Favorite color of cars				
Number of hours worked weekly				

1.2 Scatter Plots

Here you'll learn how to represent data that has no definite pattern as a scatter plot and how to draw a line of best fit for the data. You'll also learn how to make predictions using a line of best fit.

You've been exercising every week and when you go for your next doctor's visit the doctor says that the reading for your resting heart rate has changed. You start taking your own resting heart rate once a week on Mondays and relate it to the numbers of hours per week you've been exercising. How would you represent this data? Do you expect to see a correlation between the number of hours you exercise per week and your resting heart rate? How would you know if there is a correlation?

Watch This

First watch this video to learn about scatter plots.

MEDIA

Click image to the left for more content.

CK-12 Foundation: Chapter7ScatterPlotsA

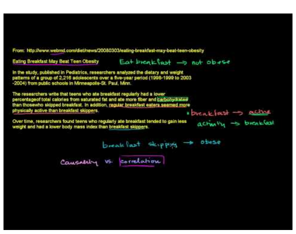
Then watch this video to see some examples.

MEDIA

Click image to the left for more content.

CK-12 Foundation: Chapter7ScatterPlotsB

Watch this video for more help.



MEDIA

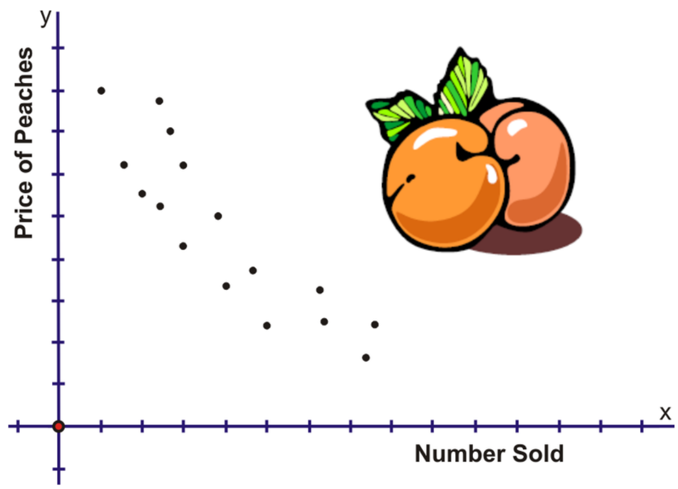
Click image to the left for more content.

Khan Academy Correlation and Causality

Guidance

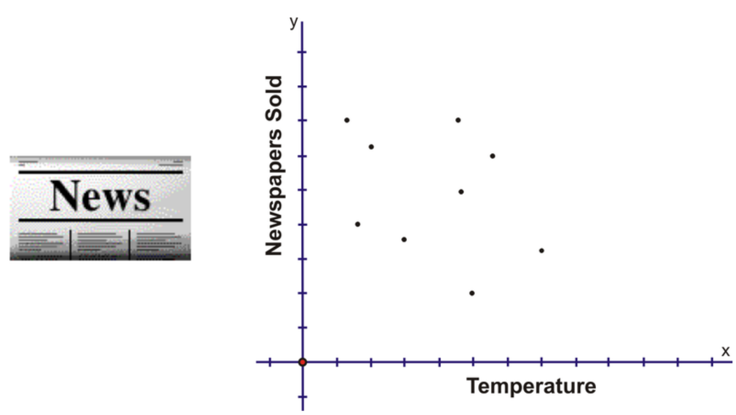
Often, when real-world data is plotted, the result is a linear pattern. The general direction of the data can be seen, but the data points do not all fall on a line. This type of graph is called a scatter plot. A **scatter plot** is often used to investigate whether or not there is a relationship or connection between 2 sets of data. The data is plotted on a graph such that one quantity is plotted on the x-axis and one quantity is plotted on the y-axis. The quantity that is plotted on the x-axis is the independent variable, and the quantity that is plotted on the y-axis is the dependent variable. If a relationship does exist between the 2 sets of data, it will be easy to see if the data is plotted on a scatter plot.

The following scatter plot shows the price of peaches and the number sold:



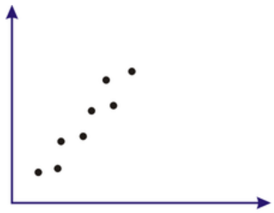
The connection is obvious—when the price of peaches was high, the sales were low, but when the price was low, the sales were high. Remember that correlation does not imply causation. There is a relationship between the number of peaches sold and the price of peaches, but the data does not determine cause and effect. Other variables such as weather, number of orchards, etc., could affect the price of peaches. The determination that one thing causes another requires a controlled experiment. For more information, please be sure to watch the last video above, Khan Academy Correlation and Causality.

The following scatter plot shows the sales of a weekly newspaper and the temperature:

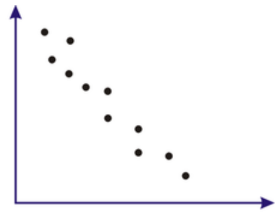


There is no connection between the number of newspapers sold and the temperature.

Another term used to describe 2 sets of data that have a connection or a relationship is **correlation**. The correlation between 2 sets of data can be positive or negative, and it can be strong or weak. The following scatter plots will help to enhance this concept.


Strong Positive Correlation

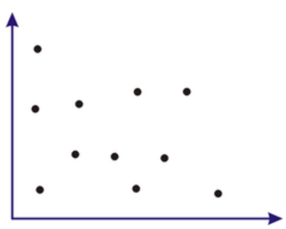

Weak Positive Correlation

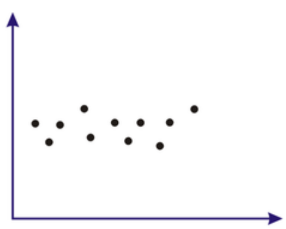

Strong Negative Correlation

If you look at the 2 sketches that represent a positive correlation, you will notice that the points are around a line that slopes upward to the right. When the correlation is negative, the line slopes downward to the right. The 2 sketches that show a strong correlation have points that are bunched together and appear to be close to a line that is in the middle of the points. When the correlation is weak, the points are more scattered and not as concentrated.

When correlation exists on a scatter plot, a line of best fit can be drawn on the graph. The **line of best fit** must be drawn so that the sums of the distances to the points on either side of the line are approximately equal and such that there are an equal number of points above and below the line. Using a clear plastic ruler makes it easier to meet all of these conditions when drawing the line. Another useful tool is a stick of spaghetti, since it can be easily rolled and moved on the graph until you are satisfied with its location. The edge of the spaghetti can be traced to produce the line of best fit. A line of best fit can be used to make estimations from the graph, but you must remember that the line of best fit is simply a sketch of where the line should appear on the graph. As a result, any values that you choose from this line are not very accurate–the values are more of a ballpark figure.

In the sales of newspapers and the temperature, there was no connection between the 2 data sets. The following sketches represent some other possible outcomes when there is no correlation between data sets:


No Correlation


No Correlation


No Correlation

An exponential function or quadratic function can also be a good fit for the data.

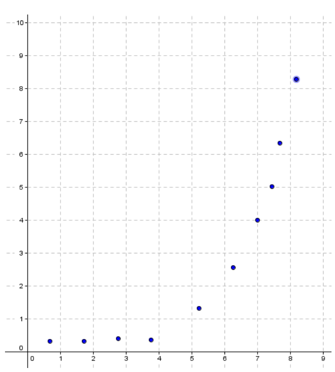


FIGURE 1.1

Exponential Model

Quadratic Model

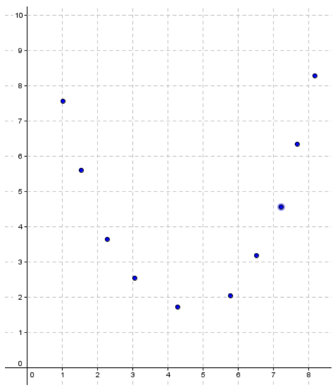
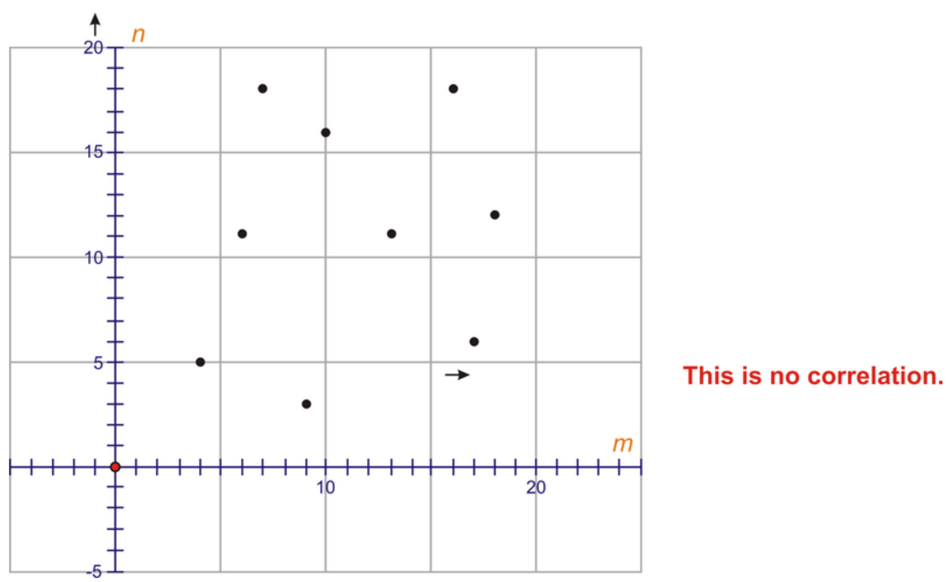


FIGURE 1.2

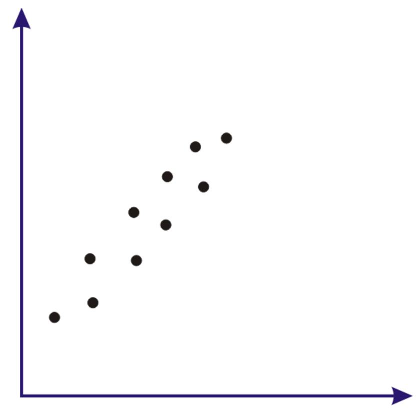
Plot the following points on a scatter plot, with m as the independent variable and n as the dependent variable. Number both axes from 0 to 20. If a correlation exists between the values of m and n, describe the correlation (strong negative, weak positive, etc.).

m	4	9	13	16	17	6	7	18	10
n	5	3	11	18	6	11	18	12	16



Example B

Describe the correlation, if any, in the following scatter plot:



In the above scatter plot, there is a strong positive correlation.

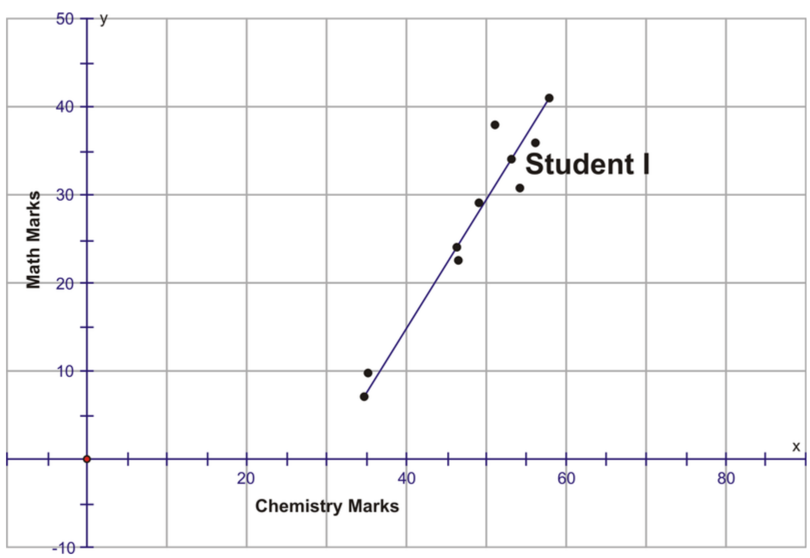
Example C

The following table consists of the marks achieved by 9 students on chemistry and math tests:

TABLE 1.6:

Student	A	B	C	D	E	F	G	H	I
Chemistry Marks	49	46	35	58	51	56	54	46	53
Math Marks	29	23	10	41	38	36	31	24	?

Plot the above marks on scatter plot, with the chemistry marks on the x-axis and the math marks on the y-axis. Draw a line of best fit, and use this line to estimate the mark that Student I would have made in math had he or she taken the test.



If Student I had taken the math test, his or her mark would have been between 32 and 37.

Points to Consider

- Can the equation for the line of best fit be used to calculate values?

- Is any other graphical representation of data used for estimations?

Guided Practice

The following table represents the sales of Volkswagen Beetles in Iowa between 1994 and 2003:

TABLE 1.7:

Year	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Beetles Sold	50	60	55	50	70	65	75	65	80	90

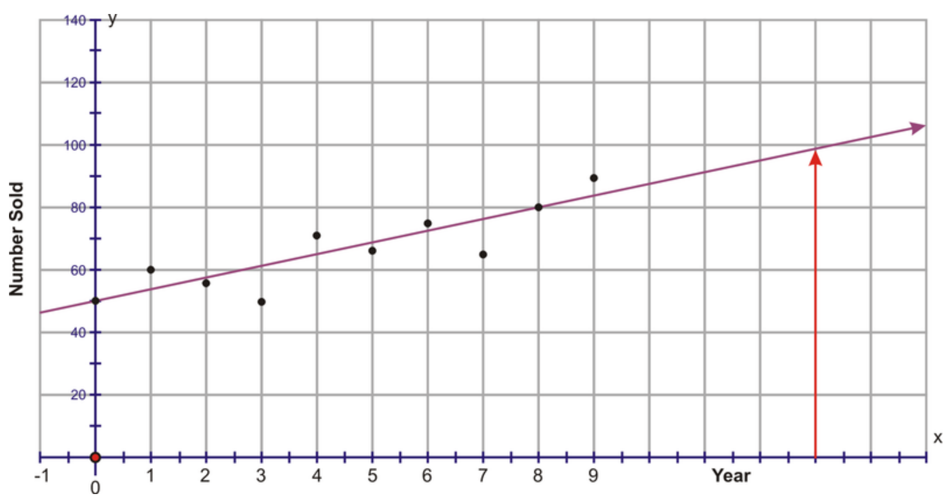
(a) Create a scatter plot and draw the line of best fit for the data. Hint: Let 0 = 1994, 1 = 1995, etc.

(b) Use the graph to predict the number of Beetles that will be sold in Iowa in the year 2007.

(c) Describe the correlation for the above graph.

Answer:

a.



b. The year 2007 would actually be the number 13 on the $x-$axis. The number of beetles sold in this year would be approximately 98 to 100.

c. The correlation of this graph is strong and positive.

Interactive Practice

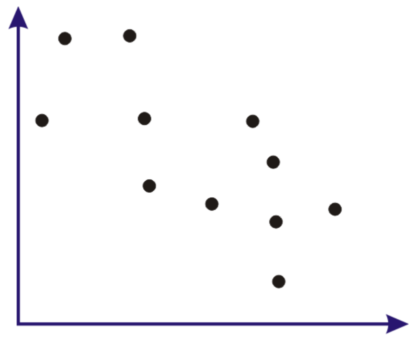
Practice

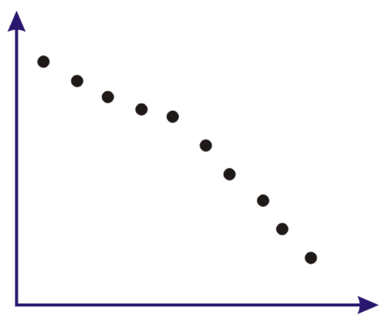
1. What is the correlation of a scatter plot that has few points that are not bunched together?
    a. strong
    b. no correlation
    c. weak
    d. negative
2. What term is used to define the connection between 2 data sets?
    a. relationship

b. scatter plot
c. correlation
d. discrete

3. Describe the correlation of each of the following graphs:


a.


b.


c.

4. Plot the following points on a scatter plot, with *m* as the independent variable and *n* as the dependent variable. Number both axes from 0 to 20. If a correlation exists between the values of *m* and *n*, describe the correlation (strong negative, weak positive, etc.).

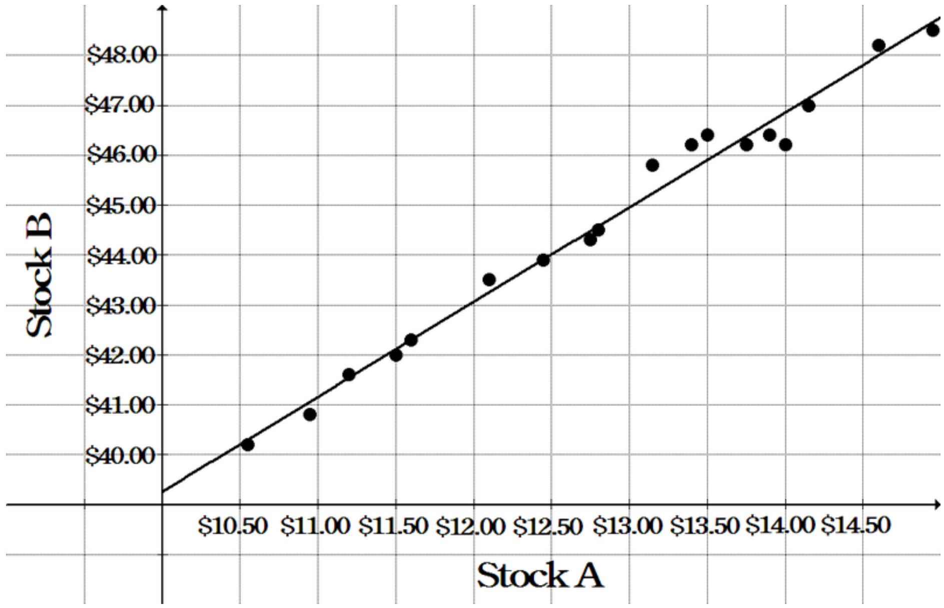
a.

m	5	14	2	10	16	4	18	2	8	11
n	6	13	4	10	15	7	16	5	8	12

b.

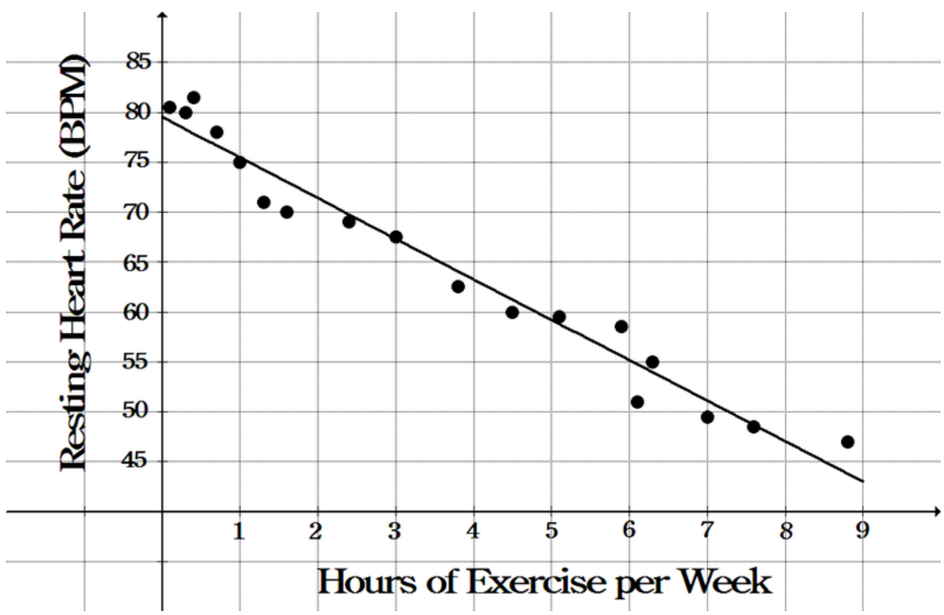
m	13	3	18	9	20	15	6	10	21	4
n	7	14	9	16	7	13	10	13	3	19

The following scatter plot shows the closing prices of 2 stocks at various points in time. A line of best fit has been drawn. Use the scatter plot to answer the following questions.



5. How would you describe the correlation between the prices of the 2 stocks?
6. If the price of stock A is $12.00, what would you expect the price of stock B to be?
7. If the price of stock B is $47.75, what would you expect the price of stock A to be?

The following scatter plot shows the hours of exercise per week and resting heart rates for various 30-year-old males. A line of best fit has been drawn. Use the scatter plot to answer the following questions.



8. How would you describe the correlation between hours of exercise per week and resting heart rate?
9. If a 30-year-old male exercises 2 hours per week, what would you expect his resting heart rate to be?
10. If a 30-year-old male has a resting heart rate of 65 beats per minute, how many hours would you expect him to exercise per week?

1.3 Scatter Plots and Linear Correlation

- Understand the concepts of bivariate data and correlation, and the use of scatterplots to display bivariate data.
- Understand when the terms 'positive', 'negative', 'strong', and 'perfect' apply to the correlation between two variables in a scatterplot graph.
- Calculate the linear correlation coefficient and coefficient of determination of bivariate data, using technology tools to assist in the calculations.
- Understand properties and common errors of correlation.

In this Concept, you will be introduced to the the concepts of bivariate data and correlation, and the use of scatterplots to display bivariate data. You will learn the terms 'positive', 'negative', 'strong', and 'perfect' and apply them to the correlation between two variables in a scatterplot graph. You will also learn to understand properties and common errors of correlation.

Watch This

For an explanation of the correlation coefficient **(13.0)**, see kbower50, The Correlation Coefficient (3:59).

MEDIA

Click image to the left for more content.

Guidance

So far we have learned how to describe distributions of a single variable and how to perform hypothesis tests concerning parameters of these distributions. But what if we notice that two variables seem to be related? We may notice that the values of two variables, such as verbal SAT score and GPA, behave in the same way and that students who have a high verbal SAT score also tend to have a high GPA (see table below). In this case, we would want to study the nature of the connection between the two variables.

TABLE 1.8: A table of verbal SAT values and GPAs for seven students.

Student	SAT Score	GPA
1	595	3.4
2	520	3.2
3	715	3.9
4	405	2.3
5	680	3.9
6	490	2.5
7	565	3.5

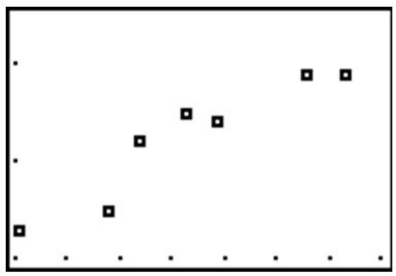
These types of studies are quite common, and we can use the concept of correlation to describe the relationship between the two variables.

Bivariate Data, Correlation Between Values, and the Use of Scatterplots

Correlation measures the relationship between bivariate data. *Bivariate data* are data sets in which each subject has two observations associated with it. In our example above, we notice that there are two observations (verbal SAT score and GPA) for each subject (in this case, a student). Can you think of other scenarios when we would use bivariate data?

If we carefully examine the data in the example above, we notice that those students with high SAT scores tend to have high GPAs, and those with low SAT scores tend to have low GPAs. In this case, there is a tendency for students to score similarly on both variables, and the performance between variables appears to be related.

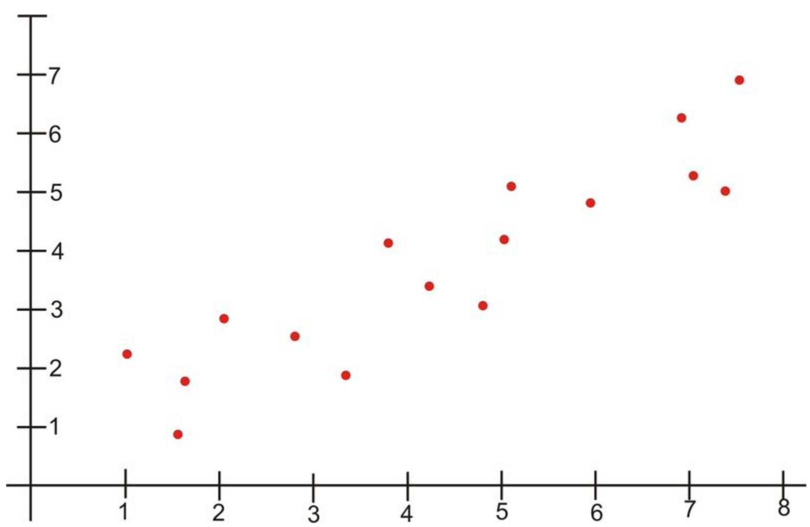
Scatterplots display these bivariate data sets and provide a visual representation of the relationship between variables. In a scatterplot, each point represents a paired measurement of two variables for a specific subject, and each subject is represented by one point on the scatterplot.



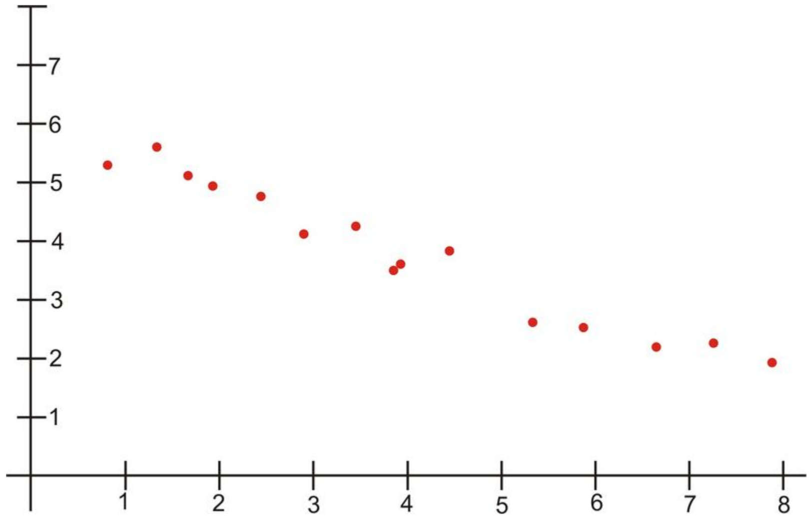
Correlation Patterns in Scatterplot Graphs

Examining a scatterplot graph allows us to obtain some idea about the relationship between two variables.

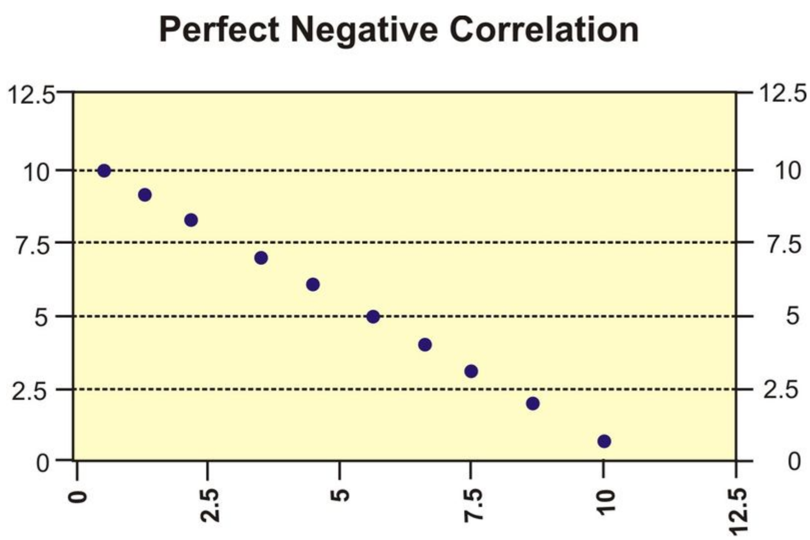
When the points on a scatterplot graph produce a lower-left-to-upper-right pattern (see below), we say that there is a *positive correlation* between the two variables. This pattern means that when the score of one observation is high, we expect the score of the other observation to be high as well, and vice versa.



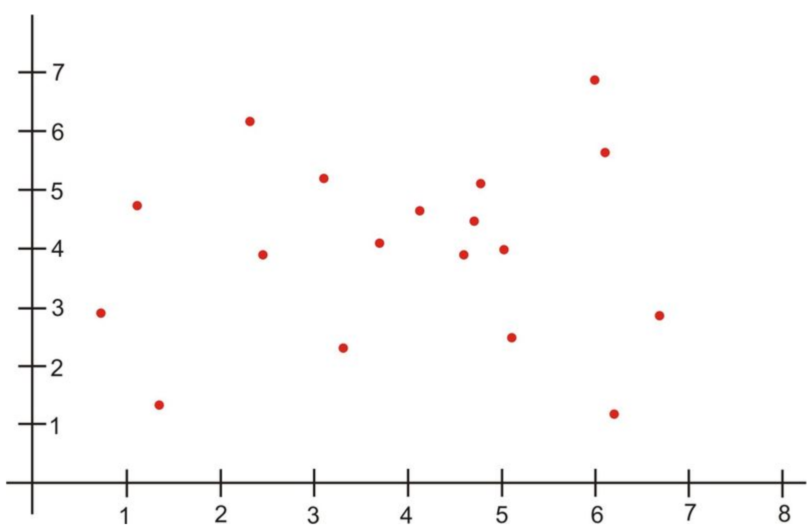
When the points on a scatterplot graph produce a upper-left-to-lower-right pattern (see below), we say that there is a *negative correlation* between the two variables. This pattern means that when the score of one observation is high, we expect the score of the other observation to be low, and vice versa.



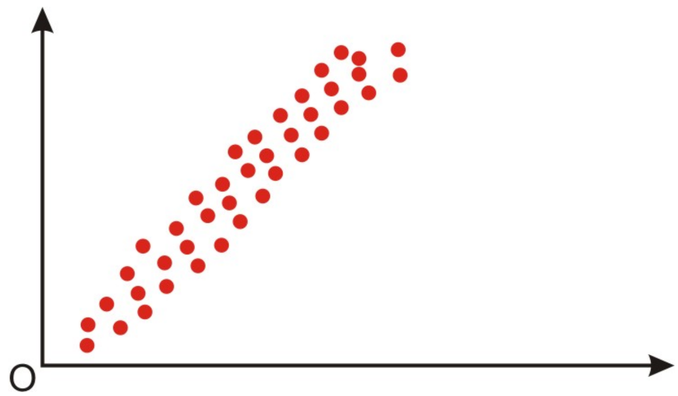
When all the points on a scatterplot lie on a straight line, you have what is called a *perfect correlation* between the two variables (see below).



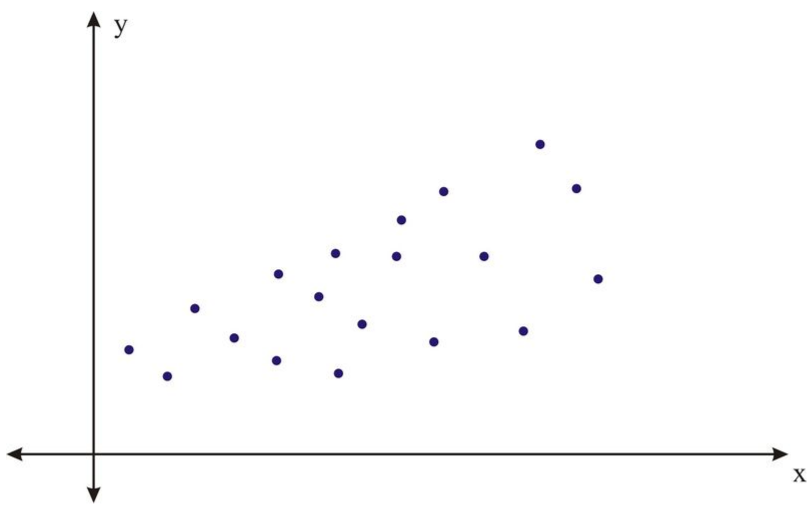
A scatterplot in which the points do not have a linear trend (either positive or negative) is called a *zero correlation* or a *near-zero correlation* (see below).



When examining scatterplots, we also want to look not only at the direction of the relationship (positive, negative, or zero), but also at the *magnitude* of the relationship. If we drew an imaginary oval around all of the points on the scatterplot, we would be able to see the extent, or the magnitude, of the relationship. If the points are close to one another and the width of the imaginary oval is small, this means that there is a strong correlation between the variables (see below).



However, if the points are far away from one another, and the imaginary oval is very wide, this means that there is a weak correlation between the variables (see below).



Correlation Coefficients

While examining scatterplots gives us some idea about the relationship between two variables, we use a statistic called the *correlation coefficient* to give us a more precise measurement of the relationship between the two variables. The correlation coefficient is an index that describes the relationship and can take on values between -1.0 and $+1.0$, with a positive correlation coefficient indicating a positive correlation and a negative correlation coefficient indicating a negative correlation.

The absolute value of the coefficient indicates the magnitude, or the strength, of the relationship. The closer the absolute value of the coefficient is to 1, the stronger the relationship. For example, a correlation coefficient of 0.20 indicates that there is a weak linear relationship between the variables, while a coefficient of -0.90 indicates that there is a strong linear relationship.

The value of a perfect positive correlation is 1.0, while the value of a perfect negative correlation is -1.0.

When there is no linear relationship between two variables, the correlation coefficient is 0. **It is important to remember that a correlation coefficient of 0 indicates that there is no *linear* relationship, but there may still**

be a strong relationship between the two variables. For example, there could be a quadratic relationship between them.

You will be calculating the correlation coefficent with technology in the next section ***Linear Regression Equations***.

On the Web

http://tinyurl.com/ylcyh88 Match the graph to its correlation.

http://tinyurl.com/y8vcm5y Guess the correlation.

http://onlinestatbook.com/stat_sim/reg_by_eye/index.html Regression by eye.

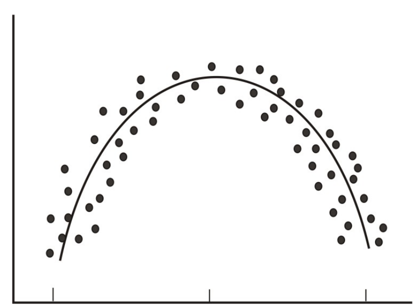
The Properties and Common Errors of Correlation

Correlation is a measure of the linear relationship between two variables-it does not necessarily state that one variable is caused by another. For example, a third variable or a combination of other things may be causing the two correlated variables to relate as they do. Therefore, it is important to remember that we are interpreting the variables and the variance not as causal, but instead as relational.

When examining correlation, there are three things that could affect our results: linearity, homogeneity of the group, and sample size.

Linearity

As mentioned, the correlation coefficient is the measure of the linear relationship between two variables. However, while many pairs of variables have a linear relationship, some do not. For example, let's consider performance anxiety. As a person's anxiety about performing increases, so does his or her performance up to a point. (We sometimes call this good stress.) However, at some point, the increase in anxiety may cause a person's performance to go down. We call these non-linear relationships *curvilinear relationships*. We can identify curvilinear relationships by examining scatterplots (see below). One may ask why curvilinear relationships pose a problem when calculating the correlation coefficient. The answer is that if we use the traditional formula to calculate these relationships, it will not be an accurate index, and we will be underestimating the relationship between the variables. If we graphed performance against anxiety, we would see that anxiety has a strong affect on performance. However, if we calculated the correlation coefficient, we would arrive at a figure around zero. Therefore, the correlation coefficient is not always the best statistic to use to understand the relationship between variables.



Homogeneity of the Group

Another error we could encounter when calculating the correlation coefficient is homogeneity of the group. When a group is homogeneous, or possesses similar characteristics, the range of scores on either or both of the variables is restricted. For example, suppose we are interested in finding out the correlation between IQ and salary. If only members of the Mensa Club (a club for people with IQs over 140) are sampled, we will most likely find a very low correlation between IQ and salary, since most members will have a consistently high IQ, but their salaries will still vary. This does not mean that there is not a relationship-it simply means that the restriction of the sample limited the magnitude of the correlation coefficient.

Sample Size

Finally, we should consider sample size. One may assume that the number of observations used in the calculation of the correlation coefficient may influence the magnitude of the coefficient itself. However, this is not the case. Yet while the sample size does not affect the correlation coefficient, it may affect the accuracy of the relationship. The larger the sample, the more accurate of a predictor the correlation coefficient will be of the relationship between the two variables.

Example A

If a pair of variables have a strong curvilinear relationship, which of the following is true:

a. The correlation coefficient will be able to indicate that a nonlinear relationship is present.

b. A scatterplot will not be needed to indicate that a nonlinear relationship is present.

c. The correlation coefficient will not be able to indicate the relationship is nonlinear.

d. The correlation coefficient will be exactly equal to zero.

Solution:

If a pair of variables have a strong curvilinear relationship

a. False, the correlation coefficient does not indicate that a curvilinear relationship is present – only that there is no linear relationship.

b. False, a scatterplot will be needed to indicate that a nonlinear relationship is present.

c. True, the correlation coefficient will not be ale to indicate the relationship is nonlinear.

d. True, the correlation coefficient is zero when there is a strong curvilinear relationship because it is a measure of a linear relationship.

Example B

A national consumer magazine reported that the correlation between car weight and car reliability is -0.30. What does this mean?

Solution:

If the correlation between car weight and car reliability is -.30 it means that as the weight of the car goes up, the reliability of the car goes down. This is not a perfect linear relationship since the absolute value of the correlation coefficient is only .30.

Vocabulary

Bivariate data are data sets with two observations that are assigned to the same subject.

Correlation measures the direction and magnitude of the linear relationship between bivariate data.

When examining **scatterplot graphs**, we can determine if correlations are positive, negative, perfect, or zero. A correlation is strong when the points in the scatterplot lie generally along a straight line.

The **correlation coefficient** is a precise measurement of the relationship between the two variables. This index can take on values between and including -1.0 and $+1.0$.

When calculating the correlation coefficient, there are several things that could affect our computation, including **curvilinear relationships**, **homogeneity** of the group, and the size of the group.

Practice

1. Give 2 scenarios or research questions where you would use bivariate data sets.
2. In the space below, draw and label four scatterplot graphs. One should show:
   a. a positive correlation
   b. a negative correlation
   c. a perfect correlation
   d. a zero correlation
3. In the space below, draw and label two scatterplot graphs. One should show:
   a. a weak correlation
   b. a strong correlation.
4. What does the correlation coefficient measure?
5. The following observations were taken for five students measuring grade and reading level.

TABLE 1.9: A table of grade and reading level for five students.

Student Number	Grade	Reading Level
1	2	6
2	6	14
3	5	12
4	4	10
5	1	4

a. Draw a scatterplot for these data. What type of relationship does this correlation have

6. What are the three factors that we should be aware of that affect the magnitude and accuracy of the Pearson correlation coefficient?

7. For each of the following pairs of variables, is there likely to be a positive association, a negative association, or no association. Explain.

a. Amount of alcohol consumed and result of a breath test.

b. Weight and grade point average for high school students.

c. Miles of running per week and time in a marathon.

8. Identify whether a scatterplot would or would not be an appropriate visual summary of the relationship between the following variables. Explain.
   a. Blood pressure and age
   b. Region of the country and opinion about gay marriage.
   c. Verbal SAT score and math SAT score.
9. Which of the numbers 0, 0.45, -1.9, -0.4, 2.6 could not be values of the correlation coefficient. Explain.
10. Which of the following implies a stronger linear relationship +0.6 or -0.8. Explain.
11. Explain how two variables can have a 0 correlation coefficient but a perfect curved relationship.

12. Describe what a scatterplot is and explain its importance.

13. Sketch and explain the following:

15. a. A scatterplot for a set of data points for which it would be appropriate to fit a regression line.
    b. A scatterplot for a set of data points for which it is not appropriate to fit a regression line.

14. Suppose data are collected for each of several randomly selected high school students for weight, in pounds, and number of calories burned in 30 minutes of walking on a treadmill at 4 mph. How would the value of the correlation coefficient change if all of the weights were converted to ounces?

15. Each of the following contains a mistake. In each case, explain what is wrong.

15.
   a. “There is a high correlation between the gender of a worker and his income.
   b. “We found a high correlation (1.10) between a high school freshman’s rating of a movie and a high school senior’s rating of the same movie.”
   c. The correlation between planting rate and yield of potatoes was $r = .25$ bushels.”

Keywords

Bivariate data

Coefficient of determination

Correlation

Correlation coefficient

Curvlinear relationship

Near-zero correlation

Negative correlation

Perfect correlation

Positive correlation

r

r^2

Regression coefficient

Scatterplots

Zero correlation

1.4 Linear Regression Equations

Here you'll learn how to use a Texas Instruments calculator to create a scatter plot and to determine the equation of the line of best fit. You'll also learn how to determine if a linear regression equation is a good fit for the data.

Suppose you have a large database that includes the scores on physics exams and calculus exams from high school students across your state who took both tests. You want to find out whether there is a correlation between these two sets of scores. What tools could you use to find out this information in an efficient way?

Watch This

First watch this video to learn about linear regression equations.

MEDIA

Click image to the left for more content.

CK-12 Foundation: Chapter7LinearRegressionEquationsA

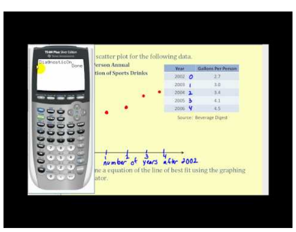
Then watch this video to see some examples.

MEDIA

Click image to the left for more content.

CK-12 Foundation: Chapter7LinearRegressionEquationsB

Watch this video for more help.



MEDIA

Click image to the left for more content.

James Sousa Linear Regression on the TI84 - Example 1

Guidance

Scatter plots and lines of best fit can also be drawn by using technology. The TI-83/84 is capable of graphing both a scatter plot and of inserting the line of best fit onto the scatter plot. The calculator is also able to find the **correlation coefficient** (r) and the **coefficient of determination** (r^2) for the linear regression equation.

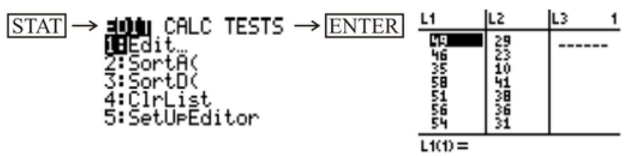
The correlation coefficient will have a value between -1 and 1. The closer the correlation coefficient is to -1 or 1, the stronger the correlation. If the correlation coefficient is negative, this implies a negative correlation, and if the correlation coefficient is positive, this implies a positive correlation. The coefficient of determination is just the correlation coefficient squared, and, therefore, it is always positive. The closer the coefficient of determination is to 1, the stronger the correlation.

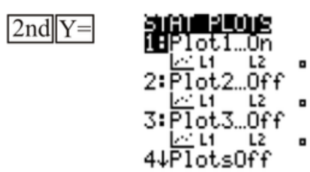
Example A

The following table consists of the marks achieved by 9 students on chemistry and math tests. Create a scatter plot for the data with your calculator.

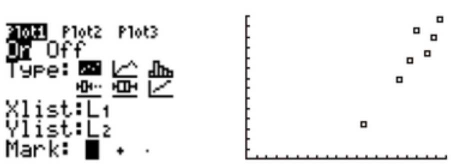
TABLE 1.10:

Student	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Chemistry Marks	49	46	35	58	51	56	54	46	53
Math Marks	29	23	10	41	38	36	31	24	?





ENTER **Scroll to the correct functions and press enter after each selection.**

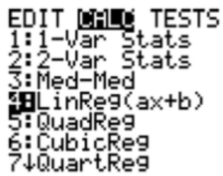


The TRACE function will give the coordinates of the points.

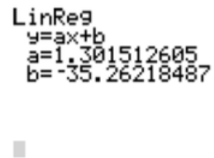
Example B

Draw a line of best fit for the data that you plotted in Example A. Use the line of best fit to calculate the predicted value for Student I's math test mark.

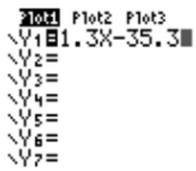
The calculator can now be used to determine a linear regression equation for the given values. The equation can be entered into the calculator, and the line will be plotted on the scatter plot.



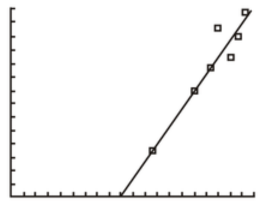
ENTER 2nd 1 , 2nd 2 ENTER

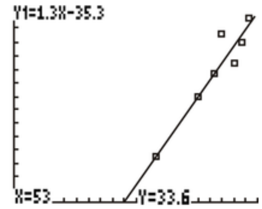


Y=



GRAPH





From the line of best fit, the calculated value for Student I's math test mark is 33.6.

Example C

Determine the correlation coefficient and the coefficient of determination for the linear regression equation that you found in Example B. Is the linear regression equation a good fit for the data?

The correlation coefficient and the coefficient of determination for the linear regression equation are found the same way that the linear regression equation is found. In other words, to find the correlation coefficient and the coefficient of determination, after entering the data into your calculator, press

STAT

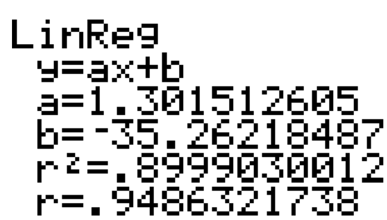
, go to the CALC menu, and choose LinReg(ax + b). After pressing

ENTER

to choose LinReg(ax + b), press

ENTER

again, and you should see the following screen:



NOTE: If you are using a TI-84 and are calculating the correlation coefficient for the first time, you need to turn your Diagnostics on. When your calculator is on, push the 2nd (blue) button and the 0 at the same time to bring up the catalog. Scroll down to Diagnostic and hit enter twice. This will allow you to see all of the statistics above.

In the example above, the equation of the line of best fit, or regression equation is $y = 1.3x - 35.3$ (rounding a and b to the nearest tenth). As the image suggests, the regression equation is in the form $y = ax + b$ where a is the slope of the line. In this case, the slope is approximately 1.3. What does this mean in the context of the data? Remember the slope can be read for every change in y, there is a change in x. In this case, for every 1.3 point gain in math, there is a 1 point gain in chemistry. The intercept, or constant term in this model is b, which is approximately -35.3. A direct interpretation of this would be if you scored a zero on the math test, the model will predict a score a -35.3 on the chemistry test. Does that seem likely? Or are there constraints on the variables in this model?

Similarly, there are limits to the prediction value of the model. Suppose you score a 5 on the math test. The model will predict a score of -28.8 on the chemistry test. Since you can't score less than a zero, this model is only accurate for scores on both tests which are greater than or equal to zero.

You can see that r, or the correlation coefficient, is equal to 0.9486321738, while r^2, or the coefficient of determination, is equal to 0.8999030012. This means that the linear regression equation is a moderately good fit, but not a great fit, for the data.

Guided Practice

The data below gives the fuel efficiency of cars with the same-sized engines when driven at various speeds.

Speed (m/h)	32	64	77	42	82	57	72
Fuel Efficiency (m/gal)	40	27	24	37	22	36	28

a. Draw a scatter plot and a line of best fit using technology. What is the equation of the line of best fit?

b. What is the correlation coefficient and the coefficient of determination of the linear regression equation? Is the linear regression equation a good fit for the data?

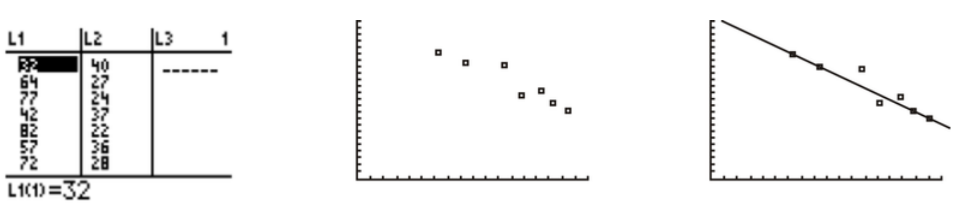
c. If a car were traveling at a speed of 47 m/h, estimate the fuel efficiency of the car.

d. If a car has a fuel efficiency of 29 m/gal, estimate the speed of the car.

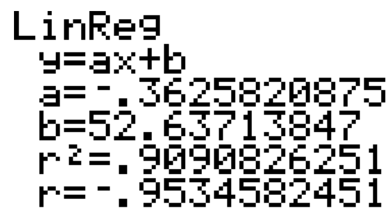
e. Interpret the slope and intercept of the model. Is this model always accurate?

Answer:

a.

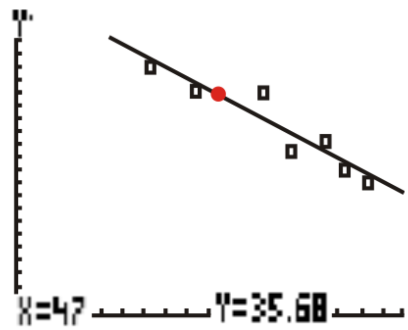


From the following screen, the equation of the line of best fit is approximately $y = -0.36x + 52.6$.



b. As can be seen in the screen in the answer to part a, the correlation coefficient is 0.9534582451, while the coefficient of determination is 0.9090826251. This means that the linear regression equation is a moderately good fit, but not a great fit, for the data.

c. Using the TI-83 to calculate the value, the fuel efficiency of a car traveling at a speed of 47 m/h would be approximately 35 m.



d. From the calculator, the equation of the line of best fit is approximately $y = -0.36x + 52.6$, where y represents the fuel efficiency of the car and x represents the speed of the car.

Using this equation:

$$\begin{aligned}
y &= -0.36x + 52.6\\
29 &= -0.36x + 52.6\\
29 - 52.6 &= -0.36x + 52.6 - 52.6\\
\frac{-23.6}{-0.36} &= \frac{-0.36x}{-0.36}\\
65.6 \text{ m/h} &= x
\end{aligned}$$

The speed of the car would be approximately 65.6 miles per hour.

e. Slope: The fuel efficiency decreases by -0.36 miles per gallon for every 1 mile per hour increase in speed. Intercept (constant term): If the car is still, the fuel efficiency would be 52.6 miles per gallon (which in this model would not make sense). The accuracy of this model is in accurate when the car is still. In addition, the model may also be inaccurate when the car is moving very slow. Generally, when you are going very slow (such as in rush hour traffic) your fuel efficiency is lower. More data would be necessary to truly predict fuel efficiency at lower miles per hour.

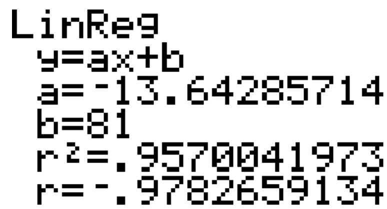
Interactive Practice

Practice

1. Which of the following calculations will create the line of best fit on the TI-83?

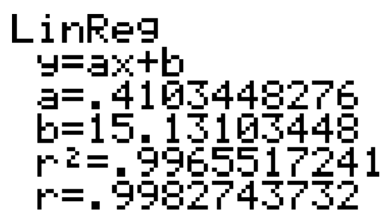
a. quadratic regression
b. cubic regression
c. exponential regression
d. linear regression $(ax+b)$

The linear regression below was performed on a data set with a TI calculator. Use the information shown on the screen to answer the following questions:



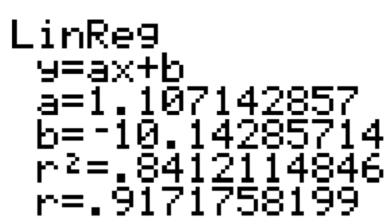
2. What is the linear regression equation?
3. What is the correlation coefficient and the coefficient of determination? Is the linear regression equation a good fit for the data?
4. According to the linear regression equation, what would be the approximate value of y when $x = 3$?

The linear regression below was performed on a data set with a TI calculator. Use the information shown on the screen to answer the following questions:



5. What is the linear regression equation?
6. What is the correlation coefficient and the coefficient of determination? Is the linear regression equation a good fit for the data?
7. According to the linear regression equation, what would be the approximate value of y when $x = 10$?

The linear regression below was performed on a data set with a TI calculator. Use the information shown on the screen to answer the following questions:



8. What is the linear regression equation?
9. What is the correlation coefficient and the coefficient of determination? Is the linear regression equation a good fit for the data?
10. According to the linear regression equation, what would be the approximate value of x when $y = 8$?
11. Lisa lights a candle and records its height in inches very hour. The results recorded as (time, height) are (0, 15), (1, 13.2), (2, 11), (3, 9.5), (4, 7.7), and (5, 6).
    (a) Express the candle's height (h) as a function of time (t) and state the meaning of the slope and the intercept in terms of the burning candle.
    (b) Draw a scatter plot and a line of best fit using technology. What is the equation of the line of best fit?
    (c) What is the correlation coefficient of the linear regression equation? Is the linear regression equation a good fit for the data?
    (d) If the candle burned for 8 hours, estimate the height of the candle.
    (e) If the candle is 4.5 inches tall, estimate the number of hours it has been burning (to the nearest hour).
    (f) Are there limitations to the accuracy of this model?

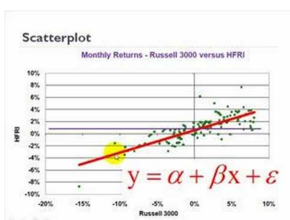
1.5 Least-Squares Regression

- Calculate and graph a regression line.
- Predict values using bivariate data plotted on a scatterplot.
- Understand outliers and influential points.
- Perform transformations to achieve linearity.
- Calculate residuals and understand the least-squares property and its relation to the regression equation.
- Plot residuals and test for linearity.

In this Concept, you will learn how to calculate and graph a regression line. You will use this to predict values using bivariate data plotted on a scatterplot. You will also learn to understand outliers and influential points, perform transformations to achieve linearity, calculate and test residuals, and test for linearity.

Watch This

For an introduction to what a least squares regression line represents **(12.0)**, see bionicturtledotcom, Introduction to Linear Regression (5:15).

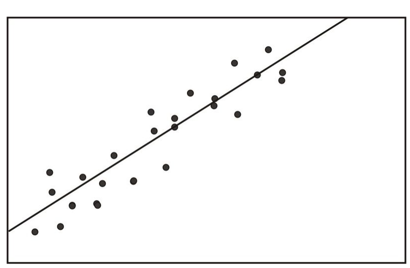


MEDIA
Click image to the left for more content.

Guidance

In a previous Concept, we learned about the concept of correlation, which we defined as the measure of the linear relationship between two variables. As a reminder, when we have a strong positive correlation, we can expect that if the score on one variable is high, the score on the other variable will also most likely be high. With correlation, we are able to roughly predict the score of one variable when we have the other. Prediction is simply the process of estimating scores of one variable based on the scores of another variable.

In a previous Concept, we illustrated the concept of correlation through scatterplot graphs. We saw that when variables were correlated, the points on a scatterplot graph tended to follow a straight line. If we could draw this straight line, it would, in theory, represent the change in one variable associated with the change in the other. This line is called the *least squares line*, or the *linear regression line* (see figure below).

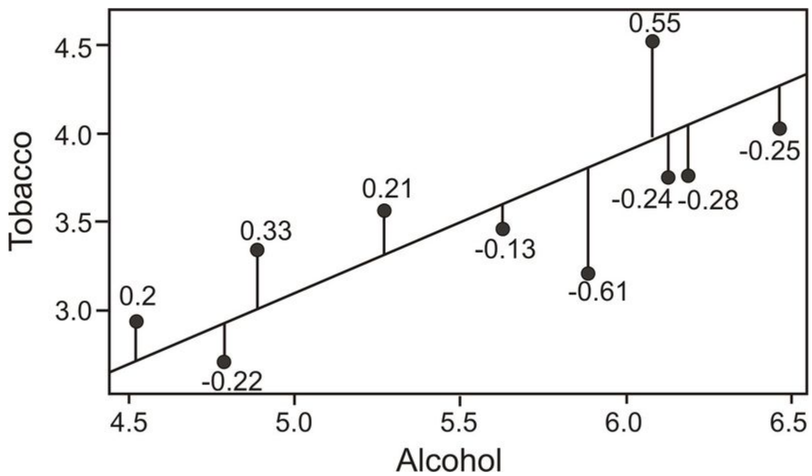


Calculating and Graphing the Regression Line

Linear regression involves using data to calculate a line that best fits that data and then using that line to predict scores. In linear regression, we use one variable (the *predictor variable*) to predict the outcome of another (the *outcome variable*, or *criterion variable*). To calculate this line, we analyze the patterns between the two variables.

We are looking for a line of best fit, and there are many ways one could define this best fit. Statisticians define this line to be the one which minimizes the sum of the squared distances from the observed data to the line.

You have used your graphing calculator to determine the line of best fit. To determine this line by hand, we want to find the change in X that will be reflected by the average change in Y. After we calculate this average change, we can apply it to any value of X to get an approximation of Y. Since the regression line is used to predict the value of Y for any given value of X, all predicted values will be located on the regression line, itself. Therefore, we try to fit the regression line to the data by having the smallest sum of squared distances possible from each of the data points to the line. In the example below, you can see the calculated distances, or **residual values**, from each of the observations to the regression line. This method of fitting the data line so that there is minimal difference between the observations and the line is called the *method of least squares*, which we will discuss further in the following sections.



As you can see, the regression line is a straight line that expresses the relationship between two variables. When predicting one score by using another, we use an equation such as the following, which is equivalent to the slope-intercept form of the equation for a straight line:

$$Y = bX + a$$

where:

Y is the score that we are trying to predict.

b is the slope of the line.

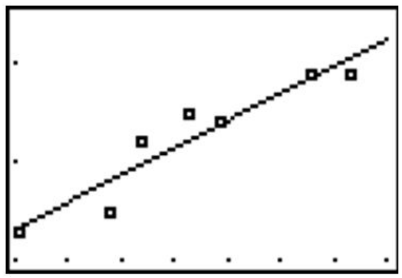
a is the y-intercept, or the value of Y when the value of X is 0.

Example A

Find the least squares line (also known as the linear regression line or the *line of best fit*) for the example measuring the verbal SAT scores and GPAs of students that was used in the previous section.

TABLE 1.11: SAT and GPA data including intermediate computations for computing a linear regression.

Student	SAT Score (X)	GPA (Y)	xy	x^2	y^2
1	595	3.4	2023	354025	11.56
2	520	3.2	1664	270400	10.24
3	715	3.9	2789	511225	15.21
4	405	2.3	932	164025	5.29
5	680	3.9	2652	462400	15.21
6	490	2.5	1225	240100	6.25
7	565	3.5	1978	319225	12.25
Sum	3970	22.7	13262	2321400	76.01



Predicting Values Using Scatterplot Data

One of the uses of a regression line is to predict values. After using your graphing calcular to determine the equation of this line, we are able to predict values by simply substituting a value of a predictor variable, X, into the regression equation and solving the equation for the outcome variable, Y. In our example above, we can predict the students' GPA's from their SAT scores by plugging in the desired values into our regression equation, $Y = 0.0056X + 0.094$.

Example B

Say that we wanted to predict the GPA for two students, one who had an SAT score of 500 and the other who had an SAT score of 600. To predict the GPA scores for these two students, we would simply plug the two values of the predictor variable into the equation and solve for Y (see below).

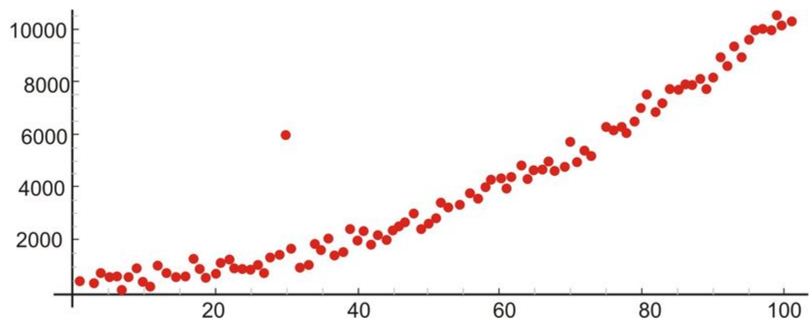
TABLE 1.12: GPA/SAT data, including predicted GPA values from the linear regression.

Student	SAT Score (X)	GPA (Y)	Predicted GPA ($\hat{Y}$)
1	595	3.4	3.4
2	520	3.2	3.0
3	715	3.9	4.1
4	405	2.3	2.3
5	680	3.9	3.9
6	490	2.5	2.8
7	565	3.5	3.2
Hypothetical	600		3.4
Hypothetical	500		2.9

As you can see, we are able to predict the value for Y for any value of X within a specified range.

Outliers and Influential Points

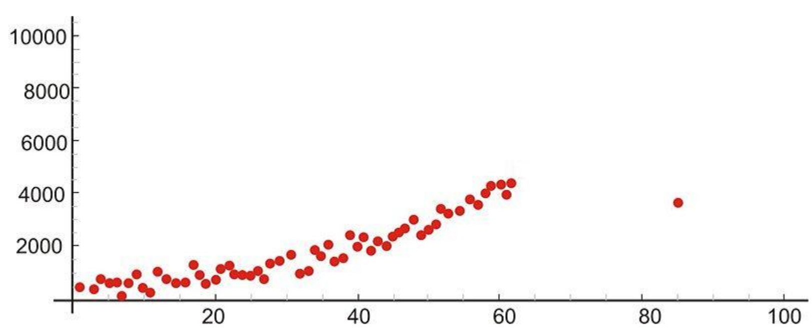
An *outlier* is an extreme observation that does not fit the general correlation or regression pattern (see figure below). In the regression setting, outliers will be far away from the regression line in the y-direction. Since it is an unusual observation, the inclusion of an outlier may affect the slope and the *y*-intercept of the regression line. When examining a scatterplot graph and calculating the regression equation, it is worth considering whether extreme observations should be included or not. In the following scatterplot, the outlier has approximate coordinates of (30, 6,000).



Let's use our example above to illustrate the effect of a single outlier. Say that we have a student who has a high GPA but who suffered from test anxiety the morning of the SAT verbal test and scored a 410. Using our original regression equation, we would expect the student to have a GPA of 2.2. But, in reality, the student has a GPA equal to 3.9. The inclusion of this value would change the slope of the regression equation from 0.0055 to 0.0032, which is quite a large difference.

There is no set rule when trying to decide whether or not to include an outlier in regression analysis. This decision depends on the sample size, how extreme the outlier is, and the normality of the distribution. For univariate data, we can use the IQR rule to determine whether or not a point is an outlier. We should consider values that are 1.5 times the inter-quartile range below the first quartile or above the third quartile as outliers. Extreme outliers are values that are 3.0 times the inter-quartile range below the first quartile or above the third quartile.

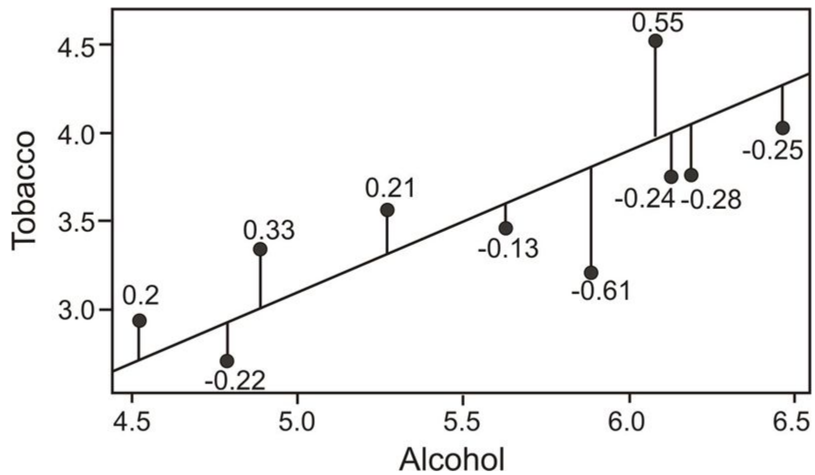
An *influential point* in regression is one whose removal would greatly impact the equation of the regression line. Usually, an influential point will be separated in the x direction from the other observations. It is possible for an outlier to be an influential point. However, there are some influential points that would not be considered outliers. These will not be far from the regression line in the y-direction (a value called a residual, discussed later) so you must look carefully for them. In the following scatterplot, the influential point has approximate coordinates of (85, 35,000).



It is important to determine whether influential points are 1) correct and 2) belong in the population. If they are not correct or do not belong, then they can be removed. If, however, an influential point is determined to indeed belong in the population and be correct, then one should consider whether other data points need to be found with similar x-values to support the data and regression line.

Calculating Residuals and Understanding their Relation to the Regression Equation

Recall that the linear regression line is the line that best fits the given data. Ideally, we would like to minimize the distances of all data points to the regression line. These distances are called the error, e, and are also known as the *residual values*. As mentioned, we fit the regression line to the data points in a scatterplot using the least-squares method. A good line will have small residuals. Notice in the figure below that the residuals are the vertical distances between the observations and the predicted values on the regression line:



To find the residual values, we subtract the predicted values from the actual values, so $e = y - \hat{y}$. Theoretically, the sum of all residual values is zero, since we are finding the line of best fit, with the predicted values as close as possible to the actual value. It does not make sense to use the sum of the residuals as an indicator of the fit, since, again, the negative and positive residuals always cancel each other out to give a sum of zero. Therefore, we try to minimize the sum of the squared residuals, or $\sum (y - \hat{y})^2$.

For additional information about residuals, including a video, click here.

Example C

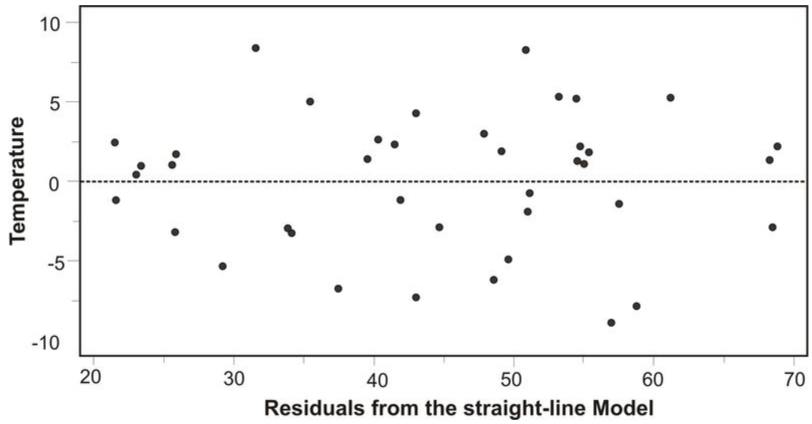
Calculate the residuals for the predicted and the actual GPA's from our sample above.

TABLE 1.13: SAT/GPA data, including residuals.

Student	SAT Score (X)	GPA (Y)	Predicted GPA $(\hat{Y})$	Residual Value	Residual Value Squared
1	595	3.4	3.4	0	0
2	520	3.2	3.0	0.2	0.04
3	715	3.9	4.1	-0.2	0.04
4	405	2.3	2.3	0	0
5	680	3.9	3.9	0	0
6	490	2.5	2.8	-0.3	0.09
7	565	3.5	3.2	0.3	0.09
$\sum (y - \hat{y})^2$					0.26

Plotting Residuals and Testing for Linearity

To test for linearity and to determine if we should drop extreme observations (or outliers) from our analysis, it is helpful to plot the residuals. When plotting, we simply plot the x-value for each observation on the x-axis and then plot the residual score on the y-axis. When examining this scatterplot, the data points should appear to have no correlation, with approximately half of the points above 0 and the other half below 0. In addition, the points should be evenly distributed along the x-axis. Below is an example of what a residual scatterplot should look like if there are no outliers and a linear relationship.



If the scatterplot of the residuals does not look similar to the one shown, we should look at the situation a bit more closely. For example, if more observations are below 0, we may have a positive outlying residual score that is skewing the distribution, and if more of the observations are above 0, we may have a negative outlying residual score. If the points are clustered close to the y-axis, we could have an x-value that is an outlier. If this occurs, we may want to consider dropping the observation to see if this would impact the plot of the residuals. If we do decide to drop the observation, we will need to recalculate the original regression line. After this recalculation, we will have a regression line that better fits a majority of the data.

Vocabulary

Prediction is simply the process of estimating scores of one variable based on the scores of another variable. We use the **least-squares regression line**, or **linear regression line**, to predict the value of a variable.

Using this regression line, we are able to use the slope, y-intercept, and the calculated **regression coefficient** to predict the scores of a variable. The predictions are represented by the variable $\hat{y}$.

The differences between the actual and the predicted values are called **residual values**. We can construct scatterplots of these residual values to examine **outliers** and **test for linearity**.

Guided Practice

Suppose a regression equation relating the average August temperature (y) and geographic latitudes (x) of 20 cities in the US is given by: $\hat{y} = 113.6 - 1.01x$

a. What is the slope of the line? Write a sentence that interprets this slope.

b. Estimate the mean August temperature for a city with latitude of 34.

c. San Francisco has a mean August temperature of 64 and latitude of 38. Use the regression equation to estimate the mean August temperature of San Francisco and determine the residual.

Solution:

a. The slope of the line is -1.01, which is the coefficient of the variable x. Since the slope is a rate of change, this slope means there is a decrease of 1.01 in temperature for each increase of 1 unit in latitude. It is a decrease in temperature because the slope is negative.

b. An estimate of the mean temperature for a city with latitude of 34 is 113.6 – 1.01(34) = 79.26 degrees.

c. San Francisco has a latitude of 38. The regression equation, therefore, estimates it mean temperature for August to be 113.6 – 1.01(38) = 75.22 degrees. The residual is 64 – 75.22 = -11.22 degrees.

Practice

1. A school nurse is interested in predicting scores on a memory test from the number of times that a student exercises per week. Below are her observations:

TABLE 1.14: A table of memory test scores compared to the number of times a student exercises per week.

Student	Exercise Per Week	Memory Test Score
1	0	15
2	2	3
3	2	12
4	1	11
5	3	5
6	1	8
7	2	15
8	0	13
9	3	2
10	3	4
11	4	2
12	1	8
13	1	10
14	1	12
15	2	8

(a) Plot this data on a scatterplot, with the x-axis representing the number of times exercising per week and the y-axis representing memory test score.

(b) Does this appear to be a linear relationship? Why or why not?

(c) What regression equation would you use to construct a linear regression model?

(d) What is the regression coefficient in this linear regression model and what does this mean in words?

(e) Calculate the regression equation for these data.

(f) Draw the regression line on the scatterplot.

(g) What is the predicted memory test score of a student who exercises 3 times per week?

(h) Calculate the residuals for each of the observations and plot these residuals on a scatterplot.

2. Suppose that the regression equation for the relationship between $y = weight$ (in pounds) and $x = height$ (in inches) for men aged 18 to 29 years old is: Average weight $y = -249 + 7x$.
   a. Estimate the average weight for men in this age group who are 68 inches tall.
   b. What is the slope of the regression line for average weight and height? Write a sentence that interprets this slope in terms of how much weight changes as height is increased by one inch.
   c. Suppose a man is 70 inches tall. Use the regression equation to predict the weight of this man.
   d. Suppose a man who is 70 inches tall weighs 193 pounds. Calculate the residual for this individual.
3. For a certain group of people, the regression line for predicting income (dollars) from education (years of schooling completed) is $y = mx + b$. The units for are ______ . The units for are __________.

4. Imagine a regression line that relates y = average systolic blood pressure to x = age. The average blood pressure of people 45 years old is 127, while for those 60 years old is 134.
   a. What is the slope of the regression line?
   b. What is the estimated average systolic blood pressure for people who are 48 years old.
5. For the following table of values calculate the regression line and then calculate $\hat{y}$ for each data point.

X	4	4	7	10	10
Y	15	11	19	21	29

6. Suppose the regression line relating verbal SAT scores and GPA is: Average GPA = 0.539 + .00365Verbal SAT
   a. Estimate the average GPA for those with verbal SAT scores of 650.
   b. Explain what the slope of 0.00365 represents in the relationship between the two variables.
   c. For two students whose verbal SAT scores differ by 50 points, what is the estimated difference in college GPAs?
   d. The lowest SAT score is 200. Does this have any useful meaning in this example? Explain.
7. A regression equation is obtained for the following set of data points: (2, 28), (4, 33), (6, 39), (6, 45), (10, 47), (12, 52). For what range of x values would it be reasonable to use the regression equation to predict the y value corresponding to the x value? Why?
8. A copy machine dealer has data on the number of copy machines, x, at each of 89 locations and the number of service calls, y, in a month at each location. Summary calculations give $\bar{x} = 8.4, s_x = 2.1, \bar{y} = 14.2, s_y = 3.8, r = 0.86$. What is the slope of the least squares regression line of number of service calls on number of copiers?
9. A study of 1,000 families gave the following results: Average height of husband is $\bar{x} = 68$ inches, $s_x = 2.7$ inches Average height of wife is $\bar{y} = 63$ inches, $s_y = 2.5$ inches, $r = .25$ Estimate the height of a wife when her husband is 73 inches tall.
10. What does it mean when a residual plot exhibits a curved pattern?
11. Hooke's Law states that, when a load (weight) is placed on a spring, length under load = constant (load) + length under no load. The following experimental results are obtained:

$Load(kg)$	$Length(cm)$
0	287.12
1	287.18
1	287.16
3	287.25
4	287.33
4	287.35
6	287.40
12	287.75

a. Find the regression equation for predicting length from load.

b. Use the equation to predict length at the following loads: 2 kg, 3 kg, 5 kg, 105 kg.

c. Estimate the length of the spring under no load.

d. Estimate the constant in Hooke's law

12. Consider the following data points: (1, 4), (2, 10), (3, 14), (4, 16) and the following possible regression lines: $\hat{y} = 3 + 3x$ and $\hat{y}1 + 4x$. By the least squares criterion which of these lines is better for this data? What is it better?
13. For ten friends (all of the same sex) determine the height and weight.
    a. Draw a scatterplot of the data with weight on the vertical axis and height on the horizontal axis. Draw a line that you believe describes the pattern.
    b. Using your calculator, compute the least squares regression line and compare the slope of this line to the slope of the line you drew in part (a).
14. Suppose a researcher is studying the relationship between two variables, x and y. For part of the analysis she computes a correlation of -.450.
    a. Explain how to interpret the reported value of the correlation.
    b. Can you tell whether the sign of the slope in the corresponding regression equation would be positive or negative? Why?
    c. Suppose the regression equation were $\hat{y} = 6.5 - 0.2x$. Interpret the slope, and show how to find the predicted y for an $x = 45$.
15. True or False:
    a. For a given set of data on two quantitative variables, the slope of the least squares regression line and the correlation must have the same sign.
    b. For a given set of data on two quantitative variables, the regression equation and the correlation do not depend on the units of measurement.
    c. Any point that is an influential observation is also an outlier, while an outlier may or may not be an influential observation.

Keywords

Criterion variable

e

Least squares line

Line of best fit

Linear regression

Linear regression line

Method of least squares

Outcome variable

Outlier

Predictor variable

r

r^2

Regression coefficient

Regression constant

Residual values

Scatterplots

CHAPTER 2 Statistics

Chapter Outline

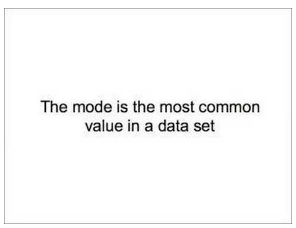
2.1 Measures of Central Tendency

- Calculate the mode, median, and mean for a set of data, and understand the differences between each measure of center.
- Identify the symbols and know the formulas for sample and population means.
- Determine the values in a data set that are outliers.
- Identify the values to be removed from a data set for an $n\%$ trimmed mean.

This Concept is an overview of some of the basic statistics used to measure the center of a set of data.

Watch This

For an explanation and examples of mean, median and mode, see keithpeterb, Mean, Mode and Median from Frequency Tables (7:06).



MEDIA
Click image to the left for more content.

Guidance

The students in a statistics class were asked to report the number of children that live in their house (including brothers and sisters temporarily away at college). The data are recorded below:

1, 3, 4, 3, 1, 2, 2, 2, 1, 2, 2, 3, 4, 5, 1, 2, 3, 2, 1, 2, 3, 6

Once data are collected, it is useful to summarize the data set by identifying a value around which the data are centered. Three commonly used measures of center are the mode, the median, and the mean.

Mode

The *mode* is defined as the most frequently occurring number in a data set. The mode is most useful in situations that involve categorical (qualitative) data that are measured at the nominal level. In the last chapter, we referred to the data with the Galapagos tortoises and noted that the variable 'Climate Type' was such a measurement. For this example, the mode is the value 'humid'.

Example A

Find the mode for the number of children per house in the data set at the beginning of the Concept.

Solution:

In this case, 2 is the mode, as it is the most frequently occurring number of children in the sample, telling us that most students in the class come from families where there are 2 children.

In this example, the mode could be a useful statistic that would tell us something about the families of statistics students in our school.

More Than One Mode

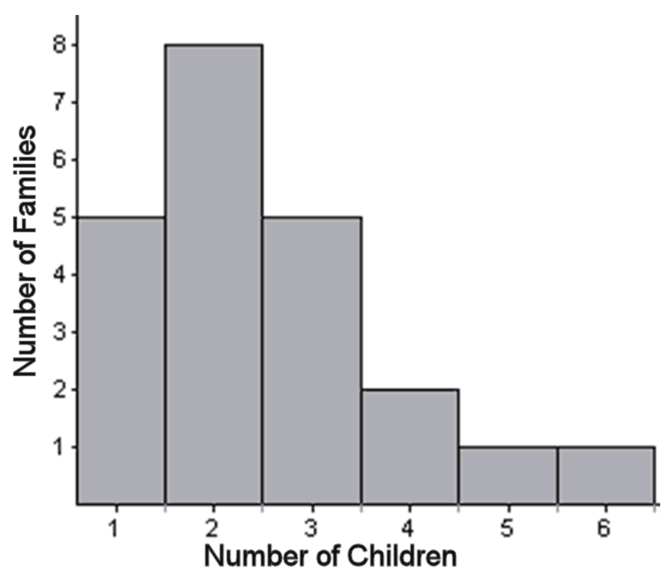
If there were seven 3-child households and seven 2-child households, we would say the data set has two modes. In other words, the data would be *bimodal*. When a data set is described as being bimodal, it is clustered about two different modes. Technically, if there were more than two, they would all be the mode. However, the more of them there are, the more trivial the mode becomes. In these cases, we would most likely search for a different statistic to describe the center of such data.

If there is an equal number of each data value, the mode is not useful in helping us understand the data, and thus, we say the data set has no mode.

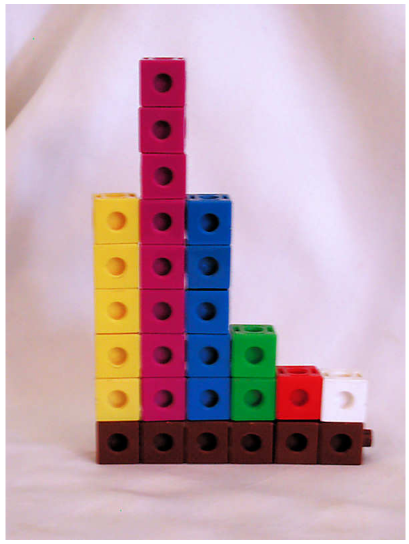
Mean

Another measure of central tendency is the arithmetic average, or *mean*. This value is calculated by adding all the data values and dividing the sum by the total number of data points. The mean is the numerical balancing point of the data set.

We can illustrate this physical interpretation of the mean. Below is a graph of the class data from the last example.



If you have snap cubes like you used to use in elementary school, you can make a physical model of the graph, using one cube to represent each student's family and a row of six cubes at the bottom to hold them together, like this:



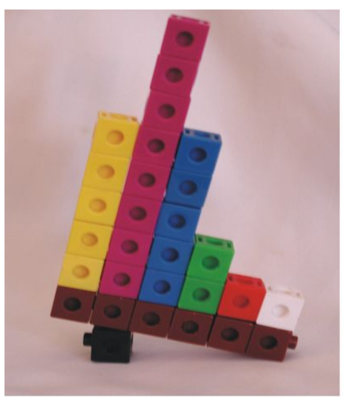
Example B

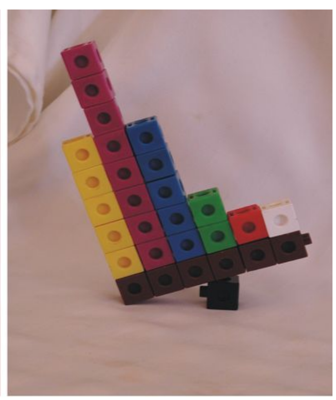
Find the mean for the number of children per house.

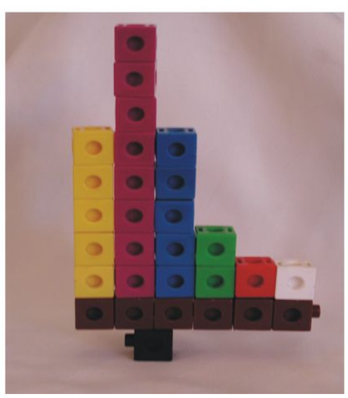
Solution:

There are 22 students in this class, and the total number of children in all of their houses is 55, so the mean of this data is $\frac{55}{22} = 2.5$.

It turns out that the model that you created balances at 2.5. In the pictures below, you can see that a block placed at 3 causes the graph to tip left, while one placed at 2 causes the graph to tip right. However, if you place the block at 2.5, it balances perfectly!







Statisticians use the symbol $\overline{x}$ to represent the mean when x is the symbol for a single measurement. Read $\overline{x}$ as "x bar."

Symbolically, the formula for the sample mean is as follows:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

where:

x_i is the i^{th} data value of the sample.

n is the sample size.

The mean of the population is denoted by the Greek letter, μ.

$\overline{x}$ is a statistic, since it is a measure of a sample, and μ is a parameter, since it is a measure of a population. $\overline{x}$ is an estimate of μ.

Median

The *median* is simply the middle number in an ordered set of data.

Suppose a student took five statistics quizzes and received the following grades:

80, 94, 75, 96, 90

To find the median, you must put the data in order. The median will be the data point that is in the middle. Placing the data in order from least to greatest yields: 75, 80, 90, 94, 96.

The middle number in this case is the third grade, or 90, so the median of this data is 90.

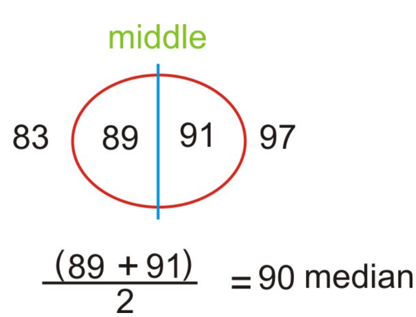
When there is an even number of numbers, no one of the data points will be in the middle. In this case, we take the average (mean) of the two middle numbers.

Example C

Consider the following quiz scores: 91, 83, 97, 89

Place them in numeric order: 83, 89, 91, 97.

The second and third numbers straddle the middle of this set. The mean of these two numbers is 90, so the median of the data is 90.



Mean vs. Median

Both the mean and the median are important and widely used measures of center. Consider the following example: Suppose you got an 85 and a 93 on your first two statistics quizzes, but then you had a really bad day and got a 14 on your next quiz!

The mean of your three grades would be 64. Which is a better measure of your performance? As you can see, the middle number in the set is an 85. That middle does not change if the lowest grade is an 84, or if the lowest grade is a 14. However, when you add the three numbers to find the mean, the sum will be much smaller if the lowest grade is a 14.

Outliers and Resistance

The mean and the median are so different in this example because there is one grade that is extremely different from the rest of the data. In statistics, we call such extreme values *outliers*. The mean is affected by the presence of an outlier; however, the median is not. A statistic that is not affected by outliers is called *resistant*. We say that the median is a resistant measure of center, and the mean is not resistant. In a sense, the median is able to resist the pull of a far away value, but the mean is drawn to such values. It cannot resist the influence of outlier values. As a result, when we have a data set that contains an outlier, it is often better to use the median to describe the center, rather than the mean.

Example D

In 2005, the CEO of Yahoo, Terry Semel, was paid almost $231,000,000 (see http://www.forbes.com/static/execpay 2005/rank.html). This is certainly not typical of what the average worker at Yahoo could expect to make. Instead of using the mean salary to describe how Yahoo pays its employees, it would be more appropriate to use the median salary of all the employees.

You will often see medians used to describe the typical value of houses in a given area, as the presence of a very few extremely large and expensive homes could make the mean appear misleadingly large.

On the Web

http://edhelper.com/statistics.htm

http://en.wikipedia.org/wiki/Arithmetic_mean

Java Applets helpful to understand the relationship between the mean and the median:

http://www.ruf.rice.edu/ lane/stat_sim/descriptive/index.html

http://www.shodor.org/interactivate/activities/PlopIt/

Vocabulary

When examining a set of data, we use **descriptive statistics** to provide information about where the data are centered:

The **mode** is a measure of the most frequently occurring number in a data set and is most useful for categorical data and data measured at the nominal level.

The **mean** and **median** are two of the most commonly used measures of center.

The **mean**, or average, is the sum of the data points divided by the total number of data points in the set. In a data set that is a sample from a population, the **sample mean** is denoted by $\bar{x}$. The **population mean** is denoted by μ.

The **median** is the numeric middle of a data set. If there are an odd number of data points, this middle value is easy to find. If there is an even number of data values, the median is the mean of the middle two values.

An **outlier** is a number that has an extreme value when compared with most of the data. The median is resistant. That is, it is not affected by the presence of outliers. The mean is not **resistant**, and therefore, the median tends to be a more appropriate measure of center to use in examples that contain outliers. Because the mean is the numerical balancing point for the data, it is an extremely important measure of center that is the basis for many other calculations and processes necessary for making useful conclusions about a set of data.

Guided Practice

The mean of 6 people in a room is 35 years. A 40- year- old person comes in. What is now the mean age of the people in the room?

Solution:

We will start by using the definition of the mean:

$\bar{x} = \frac{\Sigma x}{n}$.

Since we know the mean is 35, and that $n = 6$, so we can substitute these into the equation:

$35 = \frac{\Sigma x}{6} \Rightarrow \Sigma x = 6 \cdot 35 = 210$.

When a new person of age 40 enters the room the total becomes 210 + 40 = 250. We find the average by dividing by 7. The average age is now 35.7 years.

Practice

1. In Lois' 2nd grade class, all of the students are between 45 and 52 inches tall, except one boy, Lucas, who is 62 inches tall. Which of the following statements is true about the heights of all of the students?
   a. The mean height and the median height are about the same.
   b. The mean height is greater than the median height.
   c. The mean height is less than the median height.
   d. More information is needed to answer this question.
   e. None of the above is true.
2. Enrique has a 91, 87, and 95 for his statistics grades for the first three quarters. His mean grade for the year must be a 93 in order for him to be exempt from taking the final exam. Assuming grades are rounded following valid mathematical procedures, what is the lowest whole number grade he can get for the 4th quarter and still be exempt from taking the exam?
3. The chart below shows the data from the Galapagos tortoise preservation program with just the number of individual tortoises that were bred in captivity and reintroduced into their native habitat.

TABLE 2.1:

Island or Volcano	Number of Individuals Repatriated
Wolf	40
Darwin	0
Alcedo	0
Sierra Negra	286
Cerro Azul	357
Santa Cruz	210
Española	1293
San Cristóbal	55
Santiago	498
Pinzón	552
Pinta	0

Figure: Approximate Distribution of Giant Galapagos Tortoises in 2004 ("Estado Actual De Las Poblaciones de Tortugas Terrestres Gigantes en las Islas Galápagos," Marquez, Wiedenfeld, Snell, Fritts, MacFarland, Tapia, y Nanjoa, Scologia Aplicada, Vol. 3, Num. 1,2, pp. 98-11).

For this data, calculate each of the following:

(a) mode

(b) median

(c) mean

4. In the previous question, why is the mean signifcantly higher than the median?

5. The mean of 10 scores is 12.6. What is the sum of the scores?

6. While on vacation John drove an average of 262 miles per day for a period of 12 days. How far did John drive in total while he was on vacation?

7. Find x if 5, 9, 11, 12, 13, 14, 15 and x have a mean of 13.

8. Find a given that 3, 0, a, a, 4, a, 6, a, and 3 have a mean of 4.

9. The table below shows the results when 3 coins were tossed simultaneously 30 times. The number of tails appearing was recorded. Calculate the mode, median, and mean.

TABLE 2.2:

Number of Tails	**Number of times occurred**
3	4
2	12
1	11
0	3
Total	30

10. Compute the mean, the median and the mode for each of the following sets of numbers:

15. a. 3, 16, 3, 9, 5, 7, 11
    b. 5, 3, 3, 7, 5, 5, 16, 9, 3, 18, 11, 5, 3, 7
    c. 7, -4, 0, 12, 8, 121, -3

11. Find the mean and the median for each of the list of values:

15. a. 65, 69, 73, 77, 81, 87
    b. 11, 7, 3, 8, 101
    c. 31, 11, 41, 31

12. Find the mean and median for each of the following datasets:

15. a. 65, 66, 71, 75, 81, 85
    b. 11, 7, 1, 7, 99
    c. 31, 11, 41, 31

13. Explain why there is such a large difference between the median and the mean in the dataset of part b in the previous question

14. How do you determine which measure of center best describes a particular data set?

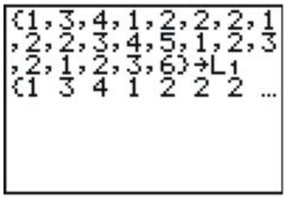
Technology Notes:

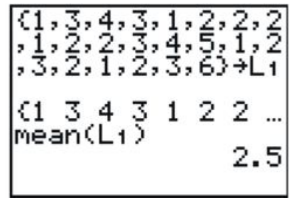
Calculating the Mean on the TI-83/84 Graphing Calculator

Step 1: Entering the data

On the home screen, press **[2ND][{]**, and then enter the following data separated by commas. When you have entered all the data, press **[2ND][}][STO][2ND][L1][ENTER]**. You will see the screen on the left below:

1, 3, 4, 3, 1, 2, 2, 2, 1, 2, 2, 3, 4, 5, 1, 2, 3, 2, 1, 2, 3, 6





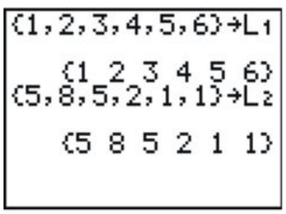
Step 2: Computing the mean

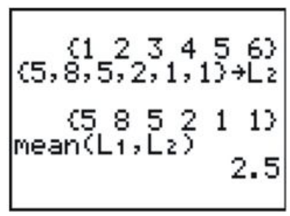
On the home screen, press **[2ND][LIST]** to enter the **LIST** menu, press the right arrow twice to go to the **MATH** menu (the middle screen above), and either arrow down and press **[ENTER]** or press **[3]** for the mean. Finally, press **[2ND][L1][)]** to insert **L1** and press **[ENTER]** (see the screen on the right above).

Calculating Weighted Means on the TI-83/84 Graphing Calculator

Use the data of the number of children in a family. In list **L1**, enter the number of children, and in list **L2**, enter the frequencies, or weights.

The data should be entered as shown in the left screen below:





Press **[2ND][STAT]** to enter the **LIST** menu, press the right arrow twice to go to the **MATH** menu (the middle screen above), and either arrow down and press **[ENTER]** or press **[3]** for the mean. Finally, press **[2ND][L1][,][2ND][L2][)][ENTER]**, and you will see the screen on the right above. Note that the mean is 2.5, as before.

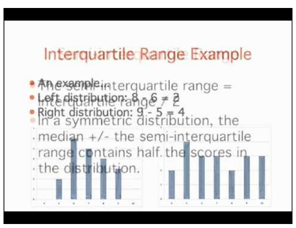
2.2 Measures of Dispersion

- Calculate the range and interquartile range.
- Calculate the standard deviation for a population and a sample, and understand its meaning.
- Distinguish between the variance and the standard deviation.
- Calculate and apply Chebyshev's Theorem to any set of data.

In this Concept, you will learn how to calculate the range, interquartile range, and the standard deviation for a population and a sample.

Watch This

For an introduction to measures of spread, or variation, see onlinestatbook, Summarizing Distributions: Measures of Variability (7:00).



MEDIA

Click image to the left for more content.

Citation: Online Statistics Education: A Multimedia Course of Study (http://onlinestatbook.com/). Project Leader: David M. Lane, Rice University.

Guidance

In the previos Concepts, we studied measures of central tendency. Another important feature that can help us understand more about a data set is the manner in which the data are distributed, or spread. Variation and dispersion are words that are also commonly used to describe this feature. There are several commonly used statistical measures of spread that we will investigate in this lesson.

Range

One measure of spread is the range. The *range* is simply the difference between the largest value (maximum) and the smallest value (minimum) in the data.

Example A

Return to the data set below:

75, 80, 90, 94, 96

The range of this data set is $96 - 75 = 21$. This is telling us the distance between the maximum and minimum values in the data set.

The range is useful because it requires very little calculation, and therefore, gives a quick and easy snapshot of how the data are spread. However, it is limited, because it only involves two values in the data set, and it is not resistant to outliers.

Interquartile Range

The *interquartile range* is the difference between the Q_3 and Q_1, and it is abbreviated IQR. Thus, $IQR = Q_3 - Q_1$. The IQR gives information about how the middle 50% of the data are spread. Fifty percent of the data values are always between Q_3 and Q_1.

5-Number Summary

The five number summary is the minimum, Q1, median, Q3, and maximum of the data set.

Example B

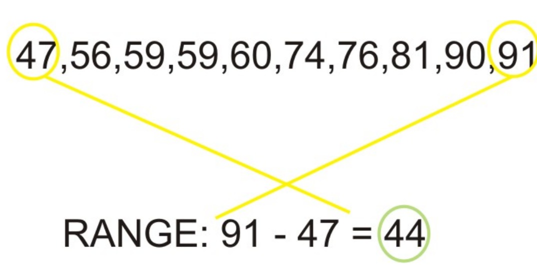
A recent study proclaimed Mobile, Alabama the wettest city in America (http://www.livescience.com/environment/070518_rainy_cities.html). The following table lists measurements of the approximate annual rainfall in Mobile over a 10 year period. Find the range and IQR for this data.

TABLE 2.3:

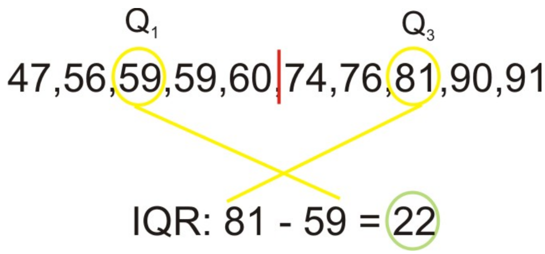
	Rainfall (inches)
1998	90
1999	56
2000	60
2001	59
2002	74
2003	76
2004	81
2005	91
2006	47
2007	59

Figure: Approximate Total Annual Rainfall, Mobile, Alabama. *Source:* http://www.cwop1353.com/CoopGaugeData.htm

First, place the data in order from smallest to largest. The range is the difference between the minimum and maximum rainfall amounts.



To find the IQR, first identify the quartiles, and then compute $Q_3 - Q_1$.



In this example, the range tells us that there is a difference of 44 inches of rainfall between the wettest and driest years in Mobile. The *IQR* shows that there is a difference of 22 inches of rainfall, even in the middle 50% of the data. It appears that Mobile experiences wide fluctuations in yearly rainfall totals, which might be explained by its position near the Gulf of Mexico and its exposure to tropical storms and hurricanes.

Note that in the example above, the bottom and the top half of the data are split evenly and you have to use an average to find the median. This makes it easy to indicate Q1 and Q3. What would happen to the median, Q1 and Q3 if we added one data point?

47, 56, ***59***, 59, 60, **74**, 76, 81, ***90***, 91, 92

The median is now 74. Now when you compute Q1 and Q3, ***you omit the 74*** when you split the data into the bottom and top halves. Q1 = 59 and Q3 = 90 and the IQR = 90 - 59 = 31

Outliers

As you know, an outlier is an extreme value. Outliers can be identified mathematically using the IQR. In order to determine if your data set has an outliers, take your IQR and multiply by 1.5. Using the original rainfall data above, the IQR is 22. Any observation that is below Q1 - 1.5 x IQR or above Q3 + 1.5 x IQR is considered an outlier. Any rainfall amount below 59 - 1.5(22) = 26 inches would be an outlier. Any rainfall amount above 81 + 1.5(22) = 114 inches would be an outlier. After examining the data, you see that we do not have any outliers in this data set.

Standard Deviation

The standard deviation is an extremely important measure of spread that is based on the mean. Recall that the mean is the numerical balancing point of the data. One way to measure how the data are spread is to look at how far away each of the values is from the mean. The difference between a data value and the mean is called the *deviation*. Written symbolically, it would be as follows:

$$\text{Deviation} = x - \bar{x}$$

Let's take the simple data set of three randomly selected individuals' shoe sizes shown below:

9.5, 11.5, 12

The mean of this data set is 11. The deviations are as follows:

TABLE 2.4: Table of Deviations

x	$x - \bar{x}$
9.5	$9.5 - 11 = -1.5$
11.5	$11.5 - 11 = 0.5$
12	$12 - 11 = 1$

Notice that if a data value is less than the mean, the deviation of that value is negative. Points that are above the mean have positive deviations.

The *standard deviation* is a measure of the typical, or average, deviation for all of the data points from the mean. However, the very property that makes the mean so special also makes it tricky to calculate a standard deviation. Because the mean is the balancing point of the data, when you add the deviations, they always sum to 0.

TABLE 2.5: Table of Deviations, Including the Sum.

Observed Data	Deviations
9.5	$9.5-11=-1.5$
11.5	$11.5-11=0.5$
12	$12-11=1$
Sum of deviations	$-1.5+0.5+1=0$

Therefore, we need all the deviations to be positive before we add them up. One way to do this would be to make them positive by taking their absolute values. This is a technique we use for a similar measure called the *mean absolute deviation*. For the standard deviation, though, we square all the deviations. The square of any real number is always positive.

TABLE 2.6:

Observed Data x	Deviation $x-\bar{x}$	$(x-\bar{x})^2$
9.5	-1.5	$(-1.5)^2=2.25$
11.5	0.5	$(0.5)^2=0.25$
12	1	1

$$\text{Sum of the squared deviations} = 2.25+0.25+1=3.5$$

We want to find the average of the squared deviations. Usually, to find an average, you divide by the number of terms in your sum. In finding the standard deviation, however, we divide by $n-1$. In this example, since $n=3$, we divide by 2. The result, which is called the *variance*, is 1.75. The variance of a sample is denoted by s^2 and is a measure of how closely the data are clustered around the mean. Because we squared the deviations before we added them, the units we were working in were also squared. To return to the original units, we must take the square root of our result: $\sqrt{1.75}\approx 1.32$. This quantity is the sample standard deviation and is denoted by s. The number indicates that in our sample, the typical data value is approximately 1.32 units away from the mean. It is a measure of how closely the data are clustered around the mean. A small standard deviation means that the data points are clustered close to the mean (which demonstrates consistency), while a large standard deviation means that the data points are spread out from the mean (which demonstrates inconsistencies).

In practice, we will not calculate the standard deviation by hand. Please scroll down to the bottom of this lesson to learn how to calculate the standard deviation with your calculator.

Sample and population standard deviation

The rainfall data was a sample of rainfall amounts take over 10 years so you would use the sample standard deviation. In most cases, you will be working with the sample standard deviation (s). In reality, we rarely have all of the data from a population. In select cases, such as when you are analyzing only your class's scores on a test, you would have a population because you have all the scores in your class. In that case, you would use the population standard deviation (σ). For more information please click here.

Example C

The following are scores for two different students on two quizzes:

Student 1: 100; 0

Student 2: 50; 50

Note that the mean score for each of these students is 50.

Student 1: Deviations: $100 - 50 = 50; \quad 0 - 50 = -50$

Squared deviations: 2500; 2500

Variance $= 5000$

Standard Deviation $= 70.7$

Student 2: Deviations: $50 - 50 = 0; \quad 50 - 50 = 0$

Squared Deviations: 0; 0

Variance $= 0$

Standard Deviation $= 0$

Student 2 has scores that are tightly clustered around the mean. In fact, the standard deviation of zero indicates that there is no variability. The student is absolutely consistent.

So, while the average of each of these students is the same (50), one of them is consistent in the work he/she does, and the other is not. This raises questions: Why did student 1 get a zero on the second quiz when he/she had a perfect paper on the first quiz? Was the student sick? Did the student forget about the quiz and not study? Or was the second quiz indicative of the work the student can do, and was the first quiz the one that was questionable? Did the student cheat on the first quiz?

There is one more question that we haven't answered regarding standard deviation, and that is, "Why $n-1$?" Dividing by $n-1$ is only necessary for the calculation of the standard deviation of a sample. When you are calculating the standard deviation of a population, you divide by N, the number of data points in your population. When you have a sample, you are not getting data for the entire population, and there is bound to be random variation due to sampling (remember that this is called sampling error).

When we claim to have the standard deviation, we are making the following statement:

"The typical distance of a point from the mean is ..."

But we might be off by a little from using a sample, so it would be better to overestimate s to represent the standard deviation.

Formulas

Sample Standard Deviation:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

where:

x_i is the i^{th} data value.

$\bar{x}$ is the mean of the sample.

n is the sample size.

Variance of a sample:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

where:

x_i is the i^{th} data value.

$\bar{x}$ is the mean of the sample.

n is the sample size.

Now that you have learned about measures of center and spread, we can take a look at center and spread visually. Please click here to see features of graphical displays.

Mean or Median? IQR or Standard Deviation?

- If data is skewed or when meaningful outliers are present, the median and 5-number summaries should be used to summarize the data.
- If the data is unimodal or symmetric, the mean and standard deviation are appropriate to describe the data.

Vocabulary

When examining a set of data, we use **descriptive statistics** to provide information about how the data are spread out.

The **range** is a measure of the difference between the smallest and largest numbers in a data set.

The **interquartile range** is the difference between the upper and lower quartiles.

Any observation that is below Q1 - 1.5 x IQR or above Q3 + 1.5 x IQR is considered an **outlier**.

A more informative measure of spread is based on the mean. We can look at how individual points vary from the mean by subtracting the mean from the data value. This is called the **deviation**. The **standard deviation** is a measure of the average deviation for the entire data set. Because the deviations always sum to zero, we find the standard deviation by adding the squared deviations. When we have the entire population, the sum of the squared deviations is divided by the population size. This value is called the **variance**. Taking the square root of the variance gives the standard deviation. For a population, the standard deviation is denoted by σ. Because a sample is prone to **random variation (sampling error)**, we adjust the sample standard deviation to make it a little larger by dividing the sum of the squared deviations by one less than the number of observations. The result of that division is the sample variance, and the square root of the sample variance is the sample standard deviation, usually notated as s.

If a distribution is **symmetric**, it can be divided at the center so that each piece is a mirror image of the other. A **unimodal distribution** has one clear peak whereas a **bimodal distribution** has two clear peaks. If there are *fewer* observations to the right of the distribution it is said to be **skewed right**, or **positively skewed**. If there are *fewer* observations to the left of the distribution it is said to be **skewed left**, or **negatively skewed**. If the data is equally spread and there are no clear peaks, the distribution is said to be **uniform** or constant. For visual illustrations of these concepts click here.

Practice

1. Following are bowling scores for two people: Luna 1 1 2 10 12 1 9 6 7 8 Chris 4 3 4 1 4 1 6 7 11 5
   a. Show that Chris and Luna have the same mean and range.
   b. Whose performance is more variable? Explain.

2. Use the rainfall data from figure 1 to answer this question.
   a. Calculate and record the sample mean:
   b. Complete the chart to calculate the standard deviation (**or use your calculator**).
   c. Are there any outliers in the data set? Explain.

d. Describe the distribution.

TABLE 2.7:

Year	Rainfall (inches)	Deviation	Squared Deviations
1998	90		
1999	56		
2000	60		
2001	59		
2002	74		
2003	76		
2004	81		
2005	91		
2006	47		
2007	59		

For 3-5, use the Galapagos Tortoise data below.

TABLE 2.8:

Island or Volcano	Number of Individuals Repatriated
Wolf	40
Darwin	0
Alcedo	0
Sierra Negra	286
Cerro Azul	357
Santa Cruz	210
Española	1293
San Cristóbal	55
Santiago	498
Pinzón	552
Pinta	0

3. Calculate the range and the IQR for this data.
4. Calculate the sample standard deviation for this data.
5. Are there any outliers in the data set above? Explain.
6. Which data set has the largest standard deviation?

5. (a) 10 10 10 10 10
   (b) 0 0 10 10 10
   (c) 0 9 10 11 20
   (d) 20 20 20 20 20

7. How do you determine which measure of spread best describes a particular data set?

8. What information does the standard deviation tell us about the specific, real data being observed?

9. What are the effects of outliers on the various measures of spread?

10. How does altering the spread of a data set affect its visual representation(s)?

Technology Notes:

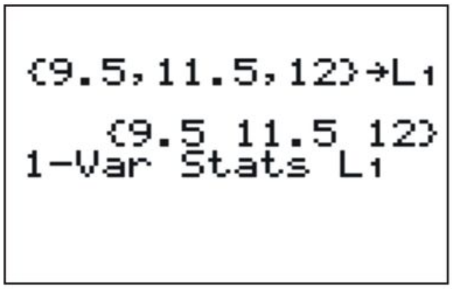
Calculating Standard Deviation on the TI-83/84 Graphing Calculator

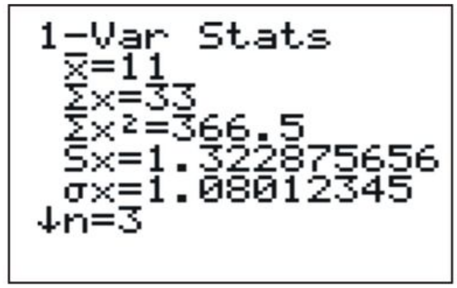
Enter the data 9.5, 11.5, 12 in list **L1** (see first screen below).

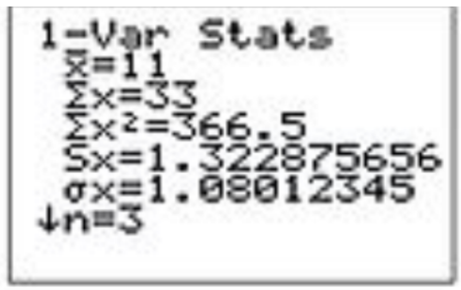
Then choose '1-Var Stats' from the **CALC** submenu of the **STAT** menu (second screen).

Enter **L1** (third screen) and press **[ENTER]** to see the fourth screen.

In the fourth screen, the symbol s_x is the sample standard deviation.







2.3 Displaying Univariate Data

- Identify and translate data sets to and from a dot plot.
- Identify and translate data sets to and from a stem-and-leaf plot.
- Identify graph distribution shapes as skewed or symmetric, and understand the basic implication of these shapes.
- Compare distributions of univariate data (shape, center, spread, and outliers).

In this Concept, we will investigate the dot plot which can be used to represent single numerical variables (univariate data). We will compare the distribution of the data, and look at the effect of outliers.

Guidance

Dot Plots

A *dot plot* is one of the simplest ways to represent numerical data. After choosing an appropriate scale on the axes, each data point is plotted as a single dot. Multiple points at the same value are stacked on top of each other using equal spacing to help convey the shape and center.

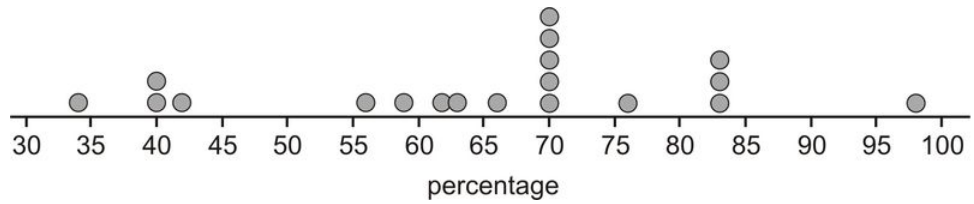
Example A

The following is a data set representing the percentage of paper packaging manufactured from recycled materials for a select group of countries.

TABLE 2.9: Percentage of the paper packaging used in a country that is recycled. Source: National Geographic, January 2008. Volume 213 No.1, pg 86-87.

Country	% of Paper Packaging Recycled
Estonia	34
New Zealand	40
Poland	40
Cyprus	42
Portugal	56
United States	59
Italy	62
Spain	63
Australia	66
Greece	70
Finland	70
Ireland	70
Netherlands	70
Sweden	70
France	76
Germany	83
Austria	83
Belgium	83
Japan	98

The dot plot for this data would look like this:



Notice that this data set is centered at a manufacturing rate for using recycled materials of between 65 and 70 percent. It is spread from 34% to 98%, and appears very roughly symmetric, perhaps even slightly skewed left. Dot plots have the advantage of showing all the data points and giving a quick and easy snapshot of the shape, center, and spread. Dot plots are not much help when there is little repetition in the data. They can also be very tedious if you are creating them by hand with large data sets, though computer software can make quick and easy work of creating dot plots from such data sets. One of the shortcomings of dot plots is that they do not show the actual values of the data. You have to read or infer them from the graph.

Vocabulary

A **dot plot** is a convenient way to represent **univariate numerical data** by plotting individual dots along a single number line to represent each value. They are especially useful in giving a quick impression of the shape, center, and spread of the data set, but are tedious to create by hand when dealing with large data sets.

Guided Practice

Here are the ages, arranged order, for the CEOs of the 60 top-ranked small companies in America in 1993 http://lib.stat.cmu.edu/DASL/Datafiles/ceodat.html

32, 33, 36, 37, 38, 40, 41, 43, 43, 44, 44, 45, 45, 45, 45,46, 46, 47, 47, 47, 48, 48, 48, 48, 49, 50, 50, 50, 50, 50, 50, 51, 51, 52, 53, 53, 53, 55, 55, 55, 56, 56, 56, 56, 57, 57, 58, 58, 59, 60, 61, 61, 61, 62, 62, 63, 69, 69, 70, 74

a) Create a dot plot for these ages.

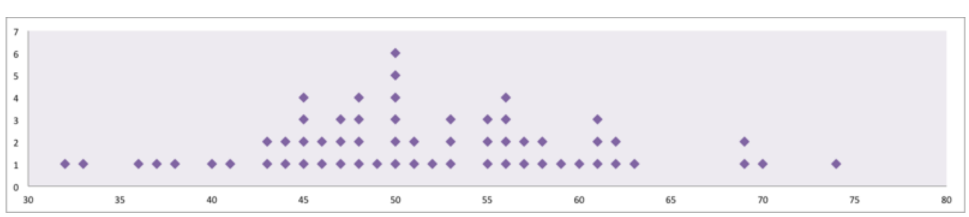
b) Describe the shape of this dataset.

c) Are there any outliers in this dataset?

Solutions:

1.

a)



b) The data set is approximately symmetric with most CEOs in their fifties.

c) There do not appear to be any outliers.

Practice

For 1-4, the following table gives the percentages of municipal waste recycled by state in the United States, including the District of Columbia, in 1998. Data was not available for Idaho or Texas.

TABLE 2.10:

State	Percentage
Alabama	23
Alaska	7
Arizona	18
Arkansas	36
California	30
Colorado	18
Connecticut	23
Delaware	31
District of Columbia	8
Florida	40
Georgia	33
Hawaii	25
Illinois	28
Indiana	23
Iowa	32
Kansas	11
Kentucky	28
Louisiana	14
Maine	41
Maryland	29
Massachusetts	33
Michigan	25
Minnesota	42
Mississippi	13
Missouri	33
Montana	5
Nebraska	27
Nevada	15
New Hampshire	25
New Jersey	45
New Mexico	12
New York	39
North Carolina	26
North Dakota	21
Ohio	19
Oklahoma	12
Oregon	28
Pennsylvania	26
Rhode Island	23
South Carolina	34
South Dakota	42
Tennessee	40
Utah	19
Vermont	30

TABLE 2.10: (continued)

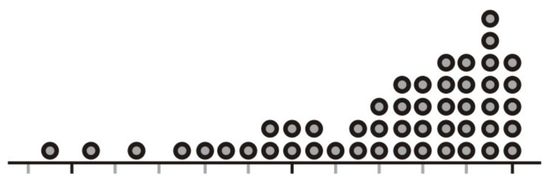
State	Percentage
Virginia	35
Washington	48
West Virginia	20
Wisconsin	36
Wyoming	5

Source: http://www.zerowasteamerica.org/MunicipalWasteManagementReport1998.htm

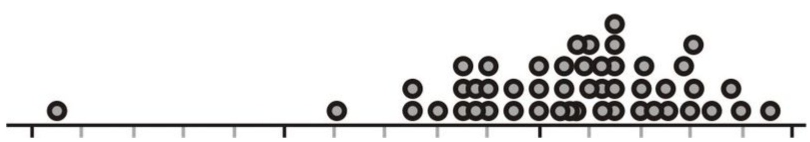
1. Create a dot plot for this data.
2. Discuss the shape, center, and spread of this distribution.

For 3-6, identify the important features of the shape of the distribution.

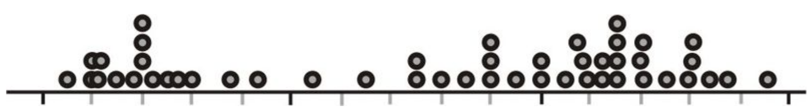
3.



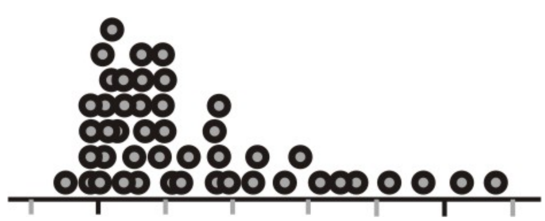
4.



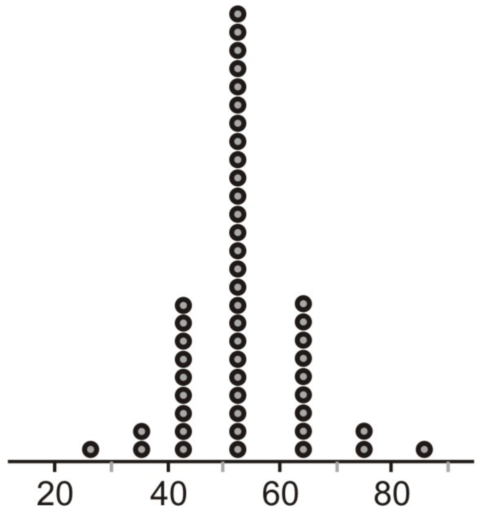
5.

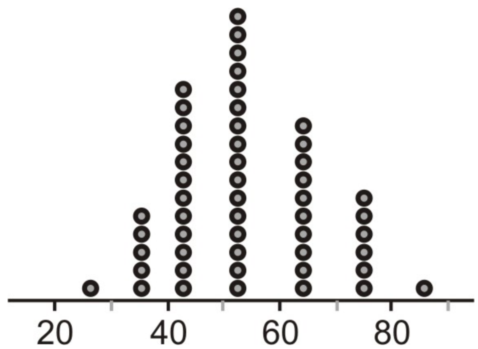


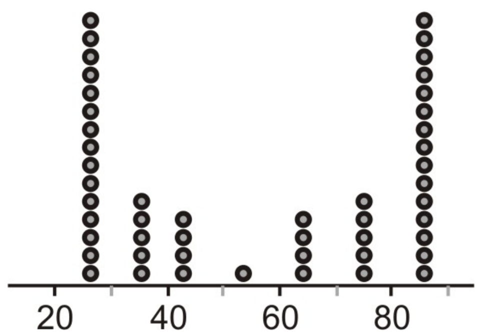
6.

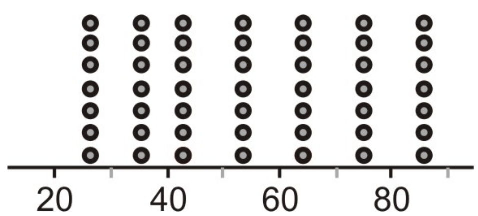


For 7-10, refer to the following dot plots:









7. Identify the overall shape of each distribution.

8. How would you characterize the center(s) of these distributions?

9. Which of these distributions has the smallest standard deviation?

10. Which of these distributions has the largest standard deviation?

11. Here are the ages, arranged order, for the CEOs of the 60 top-ranked small companies in America in 1993 http://lib.stat.cmu.edu/DASL/Datafiles/ceodat.html32, 33, 36, 37, 38, 40, 41, 43, 43, 44, 44, 45, 45, 45, 45,46, 46, 47, 47, 47, 48, 48, 48, 48, 49, 50, 50, 50, 50, 50, 50, 51, 51, 52, 53, 53, 53, 55, 55, 55, 56, 56, 56, 56, 57, 57, 58, 58, 59, 60, 61, 61, 61, 62, 62, 63, 69, 69, 70, 74

13. a. Create a dot plot for these ages.
    b. Describe the shape of this dataset.
    c. Are there any outliers in this dataset?

12. Give an example in which the same measurement taken on the same individual would be considered to be an outlier in one dataset but not in another dataset.

13. Does a five number summary provide enough information to determine if there are any outliers in the data set? Explain.

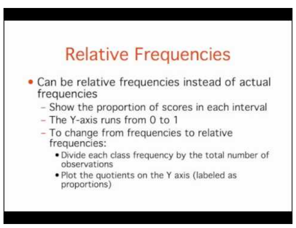
2.4 Histograms

- Read and make frequency tables for a data set.
- Identify and translate data sets to and from a histogram, a relative frequency histogram, and a frequency polygon.
- Identify histogram distribution shapes as skewed or symmetric and understand the basic implications of these shapes.
- Identify and translate data sets to and from an ogive plot (cumulative distribution function).

In this Concept, you will learn about displaying and interpreting data using a histogram.

Watch This

For a description of how to make a histogram from given data **(14.0)**, see onlinestatbook, Graphing Distributions: Histograms (6:21).



MEDIA
Click image to the left for more content.

Citation: Online Statistics Education: A Multimedia Course of Study (http://onlinestatbook.com/). Project Leader: David M. Lane, Rice University.

Guidance

The earth has seemed so large in scope for thousands of years that it is only recently that many people have begun to take seriously the idea that we live on a planet of limited and dwindling resources. This is something that residents of the Galapagos Islands are also beginning to understand. Because of its isolation and lack of resources to support large, modernized populations of humans, the problems that we face on a global level are magnified in the Galapagos. Basic human resources such as water, food, fuel, and building materials must all be brought in to the islands. More problematically, the waste products must either be disposed of in the islands, or shipped somewhere else at a prohibitive cost. As the human population grows exponentially, the Islands are confronted with the problem of what to do with all the waste. In most communities in the United States, it is easy for many to put out the trash on the street corner each week and perhaps never worry about where that trash is going. In the Galapagos, the desire to protect the fragile ecosystem from the impacts of human waste is more urgent and is resulting in a new focus on renewing, reducing, and reusing materials as much as possible. There have been recent positive efforts to encourage recycling programs.

It is not easy to bury tons of trash in solid volcanic rock. The sooner we realize that we are in the same position of limited space and that we have a need to preserve our global ecosystem, the more chance we have to save not only the uniqueness of the Galapagos Islands, but that of our own communities. All of the information in this chapter is focused around the issues and consequences of our recycling habits, or lack thereof!

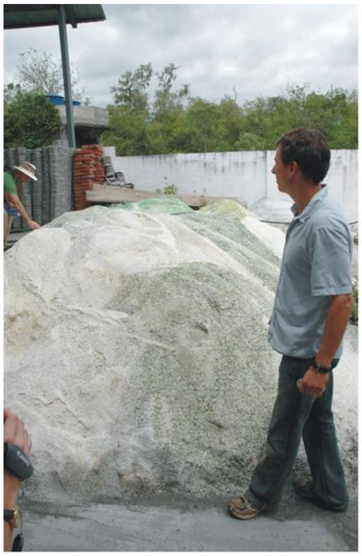


FIGURE 2.1
The Recycling Center on Santa Cruz in the Galapagos turns all the recycled glass into pavers that are used for the streets in Puerto Ayora.

Water, Water, Everywhere!

Bottled water consumption worldwide has grown, and continues to grow at a phenomenal rate. According to the Earth Policy Institute, 154 billion gallons were produced in 2004. While there are places in the world where safe water supplies are unavailable, most of the growth in consumption has been due to other reasons. The largest consumer of bottled water is the United States, which arguably could be the country with the best access to safe, convenient, and reliable sources of tap water. The large volume of toxic waste that is generated by the plastic bottles and the small fraction of the plastic that is recycled create a considerable environmental hazard. In addition, huge volumes of carbon emissions are created when these bottles are manufactured using oil and transported great distances by oil-burning vehicles.

Example A

Take an informal poll of your class. Ask each member of the class, on average, how many beverage bottles they use in a week. Once you collect this data, the first step is to organize it so it is easier to understand. A frequency table is a common starting point. *Frequency tables* simply display each value of the variable, and the number of occurrences (the frequency) of each of those values. In this example, the variable is the number of plastic beverage bottles of water consumed each week.

Consider the following raw data:

6, 4, 7, 7, 8, 5, 3, 6, 8, 6, 5, 7, 7, 5, 2, 6, 1, 3, 5, 4, 7, 4, 6, 7, 6, 6, 7, 5, 4, 6, 5, 3

Here are the correct frequencies using the imaginary data presented above:

Figure: Imaginary Class Data on Water Bottle Usage

TABLE 2.11: Completed Frequency Table for Water Bottle Data

Number of Plastic Beverage Bottles per Week	Frequency
1	1
2	1
3	3
4	4
5	6
6	8
7	7
8	2

When creating a frequency table, it is often helpful to use tally marks as a running total to avoid missing a value or over-representing another.

TABLE 2.12: Frequency table using tally marks

Number of Plastic Beverage Bottles per Week	Tally	Frequency
1	|	1
2	|	1
3	|||	3
4	||||	4
5	~~||||~~ |	6
6	~~||||~~ |||	8
7	~~||||~~ ||	7
8	||	2

The following data set shows the countries in the world that consume the most bottled water per person per year.

TABLE 2.13:

Country	Liters of Bottled Water Consumed per Person per Year
Italy	183.6
Mexico	168.5
United Arab Emirates	163.5

TABLE 2.13: (continued)

Country	Liters of Bottled Water Consumed per Person per Year
Belgium and Luxembourg	148.0
France	141.6
Spain	136.7
Germany	124.9
Lebanon	101.4
Switzerland	99.6
Cyprus	92.0
United States	90.5
Saudi Arabia	87.8
Czech Republic	87.1
Austria	82.1
Portugal	80.3

Figure: Bottled Water Consumption per Person in Leading Countries in 2004.

These data values have been measured at the ratio level. There is some flexibility required in order to create meaningful and useful categories for a frequency table. The values range from 80.3 liters to 183 liters. By examining the data, it seems appropriate for us to create our frequency table in groups of 10. We will skip the tally marks in this case, because the data values are already in numerical order, and it is easy to see how many are in each classification.

A bracket, '[' or ']', indicates that the endpoint of the interval is included in the class. A parenthesis, '(' or ')', indicates that the endpoint is not included. It is common practice in statistics to include a number that borders two classes as the larger of the two numbers in an interval. For example, $[80-90)$ means this classification includes everything from 80 and gets infinitely close to, but not equal to, 90. 90 is included in the next class, $[90-100)$.

TABLE 2.14:

Liters per Person	Frequency
$[80-90)$	4
$[90-100)$	3
$[100-110)$	1
$[110-120)$	0
$[120-130)$	1
$[130-140)$	1
$[140-150)$	2
$[150-160)$	0
$[160-170)$	2
$[170-180)$	0
$[180-190)$	1

Figure: Completed Frequency Table for World Bottled Water Consumption Data (2004)

Histograms

Once you can create a frequency table, you are ready to create our first graphical representation, called a *histogram*. Let's revisit our data about student bottled beverage habits.

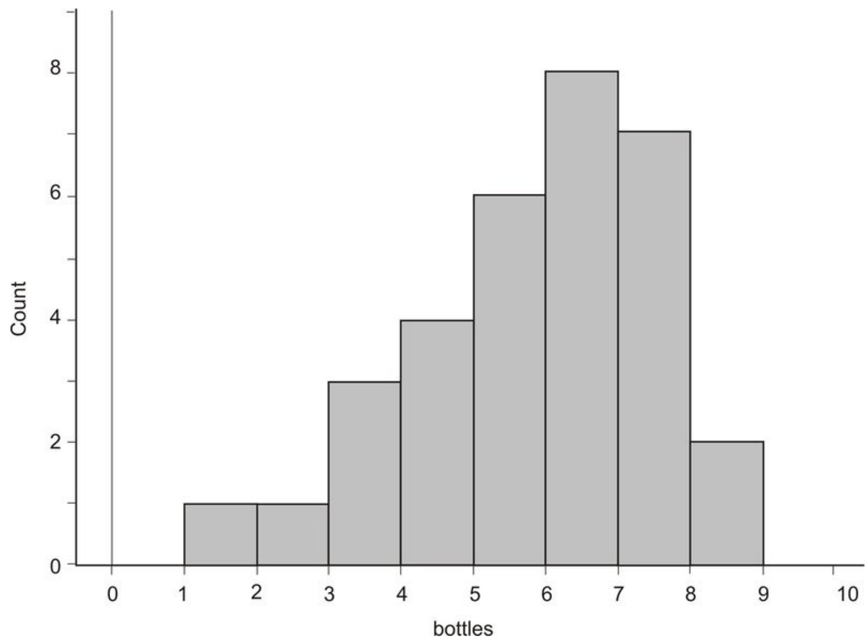
TABLE 2.15: **Completed Frequency Table for Water Bottle Data**

Number of Plastic Beverage Bottles per Week	Frequency
1	1

TABLE 2.15: (continued)

Number of Plastic Beverage Bottles per Week	Frequency
2	1
3	3
4	4
5	6
6	8
7	7
8	2

Here is the same data in a histogram:



In this case, the horizontal axis represents the variable (number of plastic bottles of water consumed), and the vertical axis is the frequency, or count. Each vertical bar represents the number of people in each class of ranges of bottles. For example, in the range of consuming $[1-2)$ bottles, there is only one person, so the height of the bar is at 1. We can see from the graph that the most common class of bottles used by people each week is the $[6-7)$ range, or six bottles per week.

A histogram is for numerical data. With histograms, the different sections are referred to as *bins*. Think of a column, or bin, as a vertical container that collects all the data for that range of values. If a value occurs on the border between two bins, it is commonly agreed that this value will go in the larger class, or the bin to the right. It is important when drawing a histogram to be certain that there are enough bins so that the last data value is included. Often this means you have to extend the horizontal axis beyond the value of the last data point. In this example, if we had stopped the graph at 8, we would have missed that data, because the 8's actually appear in the bin between 8 and 9. Very often, when you see histograms in newspapers, magazines, or online, they may instead label the midpoint of each bin. Some graphing software will also label the midpoint of each bin, unless you specify otherwise.

On the Web

http://illuminations.nctm.org/ActivityDetail.aspx?ID=78 Here you can change the bin width and explore how it effects the shape of the histogram.

Vocabulary

A **frequency table** is useful to organize data into classes according to the number of occurrences, or frequency, of each class.

Relative frequency shows the percentage of data in each class. A histogram is a graphical representation of a frequency table.

Guided Practice

There is some question as to whether caloric content listed on food products is under-reported. Look at the following table of kinds of food products (Food) and the percentage difference between measured calories and labeled calories per item (Per Item).

TABLE 2.16: Caloric Data on food items.

Food	Per item
noodles and alfredo sauce	2
cheese curls	−28
green beans	−6
mixed fruits	8
cereal	6
fig bars	−1
oatmeal raisin cookie	10
crumb cake	13
crackers	15
blue cheese dressing	−4
imperial chicken	−4
vegetable soup	−18
cheese	10
chocolate pudding	5
sausage biscuit	3
lasagna	−7
spread cheese	3
lentil soup	−0.5
pasta with shrimp and tomato sauce	−10
chocolate mousse	6
meatless sandwich	41
oatmeal cookie	46
lemon pound cake	2
banana cake	25
brownie	39
butterscotch bar	16.5
blondie	17
oat bran snack bar	28
granola bar	−3
apricot bar	14
chocolate chip cookie	34
carrot muffin	42
chinese chicken	15
gyoza	60
jelly diet candy-reds flavor	250

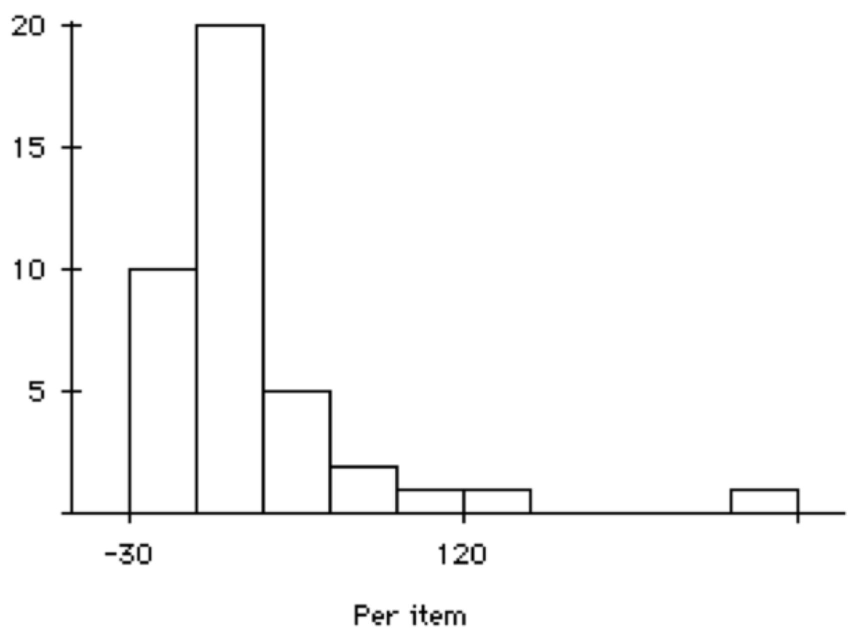
TABLE 2.16: (continued)

Food	Per item
jelly diet candy-fruit flavor	145
Florentine manicotti	6
egg foo young	80
hummus with salad	95
baba ghanoush with salad	3

Draw a histogram of the percentage difference between observed and reported calories per item.

Solution:

First, break up the percentage difference in calories per item, into intervals. By glancing at the data, it looks like using an interval length of 30 will work well, with the first interval being from -30 to zero. Count how many food items fall into each intervals, and then graph this frequency as the height on the vertical axis. For example, there are 10 food items that have a percentage difference of calories between -30 and 0, so we draw a bar with a height of 10 for that interval.



More information on this study can be found at http://lib.stat.cmu.edu/DASL/Stories/CountingCalories.html.

Practice

1. Lois was gathering data on the plastic beverage bottle consumption habits of her classmates, but she ran out of time as class was ending. When she arrived home, something had spilled in her backpack and smudged the data for the 2's. Fortunately, none of the other values was affected, and she knew there were 30 total students in the class. Complete her frequency table.

TABLE 2.17:

Number of Plastic Beverage Bottles per Week	Tally	Frequency
1	$\vert\vert$	
2		
3	$\vert\vert\vert$	
4	$\vert\vert$	
5	$\vert\vert\vert$	
6	$\cancel{\vert\vert\vert\vert}\ \vert\vert$	
7	$\cancel{\vert\vert\vert\vert}\ \vert$	
8	$\vert$	

2. The following frequency table contains exactly one data value that is a positive multiple of ten. What must that value be?

   a. 10
   b. 20
   c. 30
   d. 40
   e. There is not enough information to determine the answer.

TABLE 2.18:

Class	Frequency
$[0-5)$	4
$[5-10)$	0
$[10-15)$	2
$[15-20)$	1
$[20-25)$	0
$[25-30)$	3
$[30-35)$	0
$[35-40)$	1

3. The following table includes the data from the same group of countries from the earlier bottled water consumption example, but is for the year 1999, instead.

TABLE 2.19:

Country	Liters of Bottled Water Consumed per Person per Year
Italy	154.8
Mexico	117.0
United Arab Emirates	109.8
Belgium and Luxembourg	121.9
France	117.3
Spain	101.8
Germany	100.7
Lebanon	67.8
Switzerland	90.1
Cyprus	67.4
United States	63.6
Saudi Arabia	75.3
Czech Republic	62.1
Austria	74.6
Portugal	70.4

Figure: Bottled Water Consumption per Person in Leading Countries in 1999.

a. Create a frequency table for this data set.

b. Create the histogram for this data set.

c. How would you describe the shape of this data set?

d. Are there any outliers?

4. The following table shows the potential energy that could be saved by manufacturing each type of material using the maximum percentage of recycled materials, as opposed to using all new materials.

TABLE 2.20:

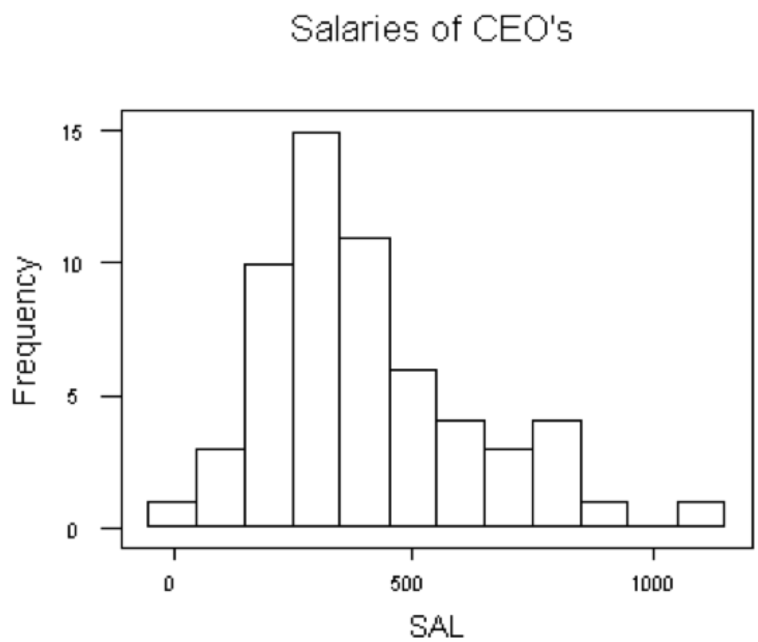
Manufactured Material	Energy Saved (millions of BTU's per ton)
Aluminum Cans	206
Copper Wire	83
Steel Cans	20
LDPE Plastics (e.g., trash bags)	56
PET Plastics (e.g., beverage bottles)	53
HDPE Plastics (e.g., household cleaner bottles)	51
Personal Computers	43
Carpet	106
Glass	2
Corrugated Cardboard	15
Newspaper	16
Phone Books	11
Magazines	11
Office Paper	10

Amount of energy saved by manufacturing different materials using the maximum percentage of recycled material as opposed to using all new material. *Source:* National Geographic, January 2008. Volume 213 No., pg 82-83.

a. Construct a frequency table. Assume a bin width of 25 million BTUs.

b. Comment on the shape, center, and spread of this distribution as it relates to the original data. (Do not actually calculate any specific statistics).

c. Are there any outliers? Explain.



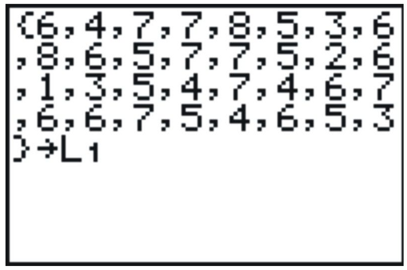
5. The figure above is a histogram of the salaries of CEOs.
    a. Are there any outliers? For any outlier, give a value for the salary and explain why you think it is an outlier.
    b. What is the salary that occurs most often? Roughly, how many CEO's report having this salary?
    c. Roughly, how many CEOs report having $500,000?
6. Forbes, November 8, 1993, "America's Best Small Companies" provided data on the salaries and age of the chief executive offers (including bonuses) of small companies. Below is a table of the age of the CEO first 60 rank companies. Create a histogram for the age of the CEO. Provide a summary of the dataset based on your histogram.

Age of CEO	Frequency
30 – 35	2
36 – 40	3
41 – 45	6
46 – 50	14
51 – 55	12
56 – 60	12
61 – 65	7
6 – 70	2
71 – 75	2

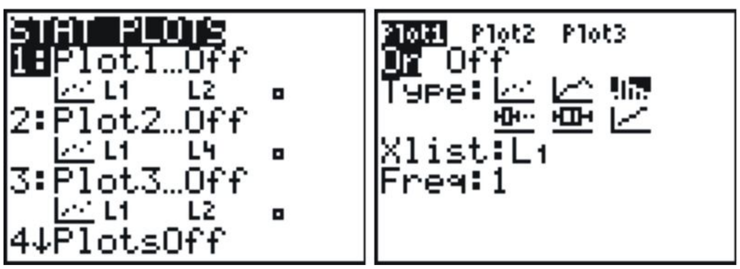
7. What effects does the shape of a data set have on the statistical measures of center and spread?
8. How do you determine the most appropriate classification to use for a frequency table or the bin width to use for a histogram?

Technology Notes: **Histograms on the TI-83/84 Graphing Calculator**

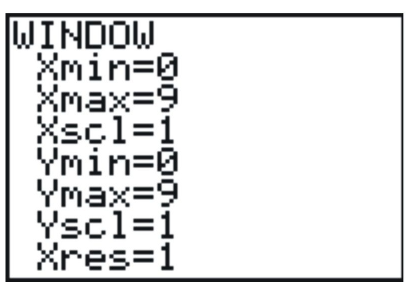
To draw a histogram on your TI-83/84 graphing calculator, you must first enter the data in a list. In the home screen, press **[2ND][{]**, and then enter the data separated by commas (see the screen below). When all the data have been entered, press **[2ND][}][STO]**, and then press **[2ND][L1][ENTER]**.



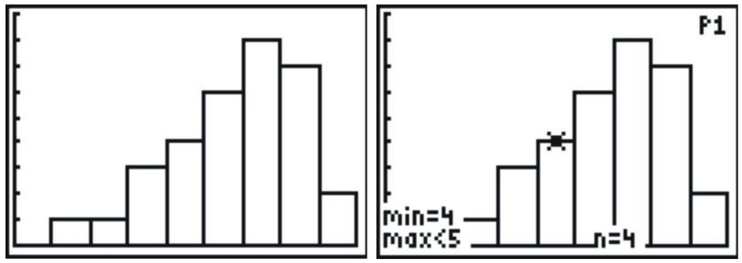
Now you are ready to plot the histogram. Press **[2ND][STAT PLOT]** to enter the **STAT-PLOTS** menu. You can plot up to three statistical plots at one time. Choose **Plot1**. Turn the plot on, change the type of plot to a histogram (see sample screen below), and choose **L1**. Enter '1' for the Freq by pressing **[2ND][A-LOCK]** to turn off alpha lock, which is normally on in this menu, because most of the time you would want to enter a variable here. An alternative would be to enter the values of the variables in **L1** and the frequencies in **L2** as we did in Chapter 1.



Finally, we need to set a window. Press **[WINDOW]** and enter an appropriate window to display the plot. In this case, 'XSCL' is what determines the bin width. Also notice that the maximum x value needs to go up to 9 to show the last bin, even though the data values stop at 8. Enter all of the values shown below.



Press **[GRAPH]** to display the histogram. If you press **[TRACE]** and then use the left or right arrows to trace along the graph, notice how the calculator uses the notation to properly represent the values in each bin.


P1
min=4
max<5
n=4

2.5 Box-and-Whisker Plots

Here you'll learn another way to graphically display a data set, called a box-and-whisker plot. You'll also learn how to interpret such displays and how to determine the effect of outliers on a data set.

What if your teacher recorded each of her student's scores on the last math test? How could she display that data in such a way that it was broken up into four distinct segments? After completing this Concept, you'll be able to make and interpret box-and-whisker plots for data such as this.

Guidance

Consider the following list of numbers: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

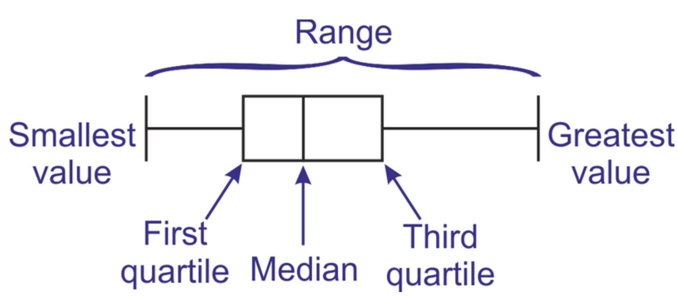
The median is the $\left(\frac{n+1}{2}\right)$ th value. There are 10 values, so the median lies halfway between the 5^{th} and the 6^{th} value. The median is therefore 5.5. This splits the list cleanly into two halves.

The lower list is: 1, 2, 3, 4, 5

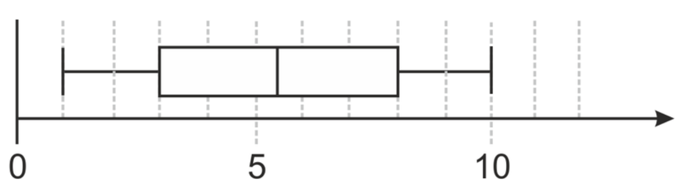
And the upper list is: 6, 7, 8, 9, 10

The median of the lower half is 3. The median of the upper half is 8. These numbers, together with the median, cut the list into four quarters. We call the division between the lower two quarters the **first quartile**. The division between the upper two quarters is the **third quartile** (the **second quartile** is, of course, the median).

A **box-and-whisker plot** is formed by placing vertical lines at five positions, corresponding to the smallest value, the first quartile, the median, the third quartile and the greatest value. (These five numbers are often referred to as the **five number summary**.) A **box** is drawn between the position of the first and third quartiles, and horizontal line segments (the **whiskers**) connect the box with the two extreme values.



The box-and-whisker plot for the integers 1 through 10 is shown below.



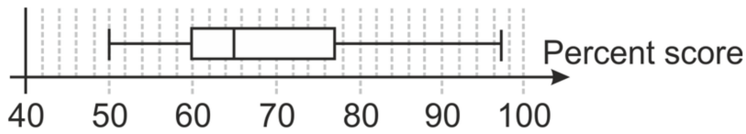
With a box-and-whisker plot, a simple measure of dispersion can be gained from the distance from the first quartile to the third quartile. This **inter-quartile range** is a measure of the spread of the middle half of the data.

Example A

Forty students took a college algebra entrance test and the results are summarized in the box-and-whisker plot below. How many students would be allowed to enroll in the class if the pass mark was set at

a) 65%

b) 60%



Solution

From the plot, we can see the following information:

Lowest score = 50%

First quartile = 60%

Median score = 65%

Third quartile = 77%

Highest score = 97%

Since the pass marks given in the question correspond with the median and the first quartile, the question is really asking how many students there are in: a) the upper half and b) the upper 3 quartiles.

a) Since there are 40 students, there are 20 in the upper half; that is, **20 students** scored above 65%.

b) Similarly, there are 30 students in the upper 3 quartiles, so **30 students** scored above 60%.

Example B

Harika is rolling 3 dice and adding the numbers together. She records the total score for each of 50 rolls, and the scores she gets are shown below. Display the data in a box-and-whisker plot, and find both the ***range*** *and the* ***inter-quartile range****.*

9, 10, 12, 13, 10, 14, 8, 10, 12, 6, 8, 11, 12, 12, 9, 11, 10, 15, 10, 8, 8, 12, 10, 14, 10, 9, 7, 5, 11, 15, 8, 9, 17, 12, 12, 13, 7, 14, 6, 17, 11, 15, 10, 13, 9, 7, 12, 13, 10, 12

Solution

First we'll put the list in order. Since there are 50 data points, $\left(\frac{n+1}{2}\right) = 26.5$, so the median will be the mean of the 25^{th} and 26^{th} values. The median will split the data into two lists of 25 values; we can write them as two distinct lists.

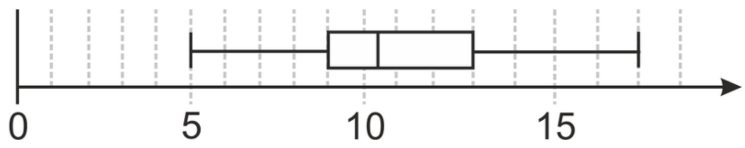
$$5,6,6,7,7,7,8,8,8,8,8,9,\boxed{9},9,9,9,10,10,10,10,10,10,10,10,\boxed{10},\boxed{11},11,$$
$$11,11,12,12,12,12,12,12,12,12,12,\boxed{13},13,13,13,14,14,14,15,15,15,17,17$$

Since each sub-list has 25 values, the first and third quartiles of the entire data set can be found from the median of each smaller list. For 25 values, $\left(\frac{n+1}{2}\right) = 13$, and so the quartiles are given by the 13^{th} value from each smaller sub-list.

From the ordered list we can see the five number summary:

- The lowest value is 5
- The first quartile is 9
- The median is 10.5
- The third quartile is 13
- The highest value is 17.

The box-and-whisker plot therefore looks like this:



The **range** is given by subtracting the smallest value from the largest value: $17 - 5 = 12$.

The **inter-quartile range** is given by subtracting the first quartile from the third quartile: $13 - 9 = 4$.

Representing Outliers in a Box-and-Whisker Plot

Box-and-whisker plots can be misleading if we don't take outliers into account. An **outlier** is a data point that does not fit well with the other data in the list. For box-and-whisker plots, we can define which points are outliers by how far they are from the *box* part of the diagram. Defining which data are outliers is somewhat arbitrary, but many books use the norm that follows. Our basic measure of distance will be the inter-quartile range (IQR).

- A **mild outlier** is a point that falls more than 1.5 times the IQR outside of the box.
- An **extreme outlier** is a point that falls more than 3 times the IQR outside of the box.

When we draw a box-and-whisker plot, we don't include the outliers in the "whisker" part of the plot; instead, we draw them as separate points.

Example C

Draw a box-and-whisker plot for the following ordered list of data:

$$1, 2, 5, 9, 10, 10, 11, 12, 13, 13, 14, 19, 25, 30$$

Solution

From the ordered list we see:

- The lowest value is 1.
- The first quartile (Q_1) is 9.
- The median is 11.5.
- The third quartile (Q_3) is 14.
- The highest value is 30.

Before we start to draw our box-and-whisker plot, we can determine the IQR:

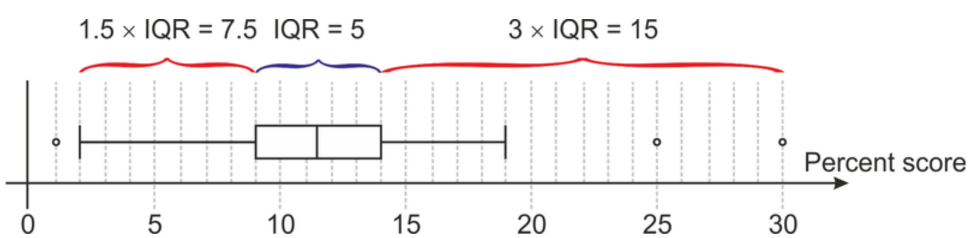
$$IQR = Q_3 - Q_1 = 14 - 9 = 5$$

Outliers are points that fall more than 1.5 times the IQR outside of the box—in other words, values that are more than 7.5 units less than 9 or greater than 14. So any values less than 1.5 or greater than 21.5 are outliers.

Looking back at the data we see:

- The value of 1 is less than 1.5, so it is a **mild outlier**.
- The value 2 is the **lowest value that falls within the included range**.
- The value 30 is greater than 21.5. In fact, it's not just more than 7.5 units outside the box, it's more than twice that far outside the box. Since it falls more than 3 times the IQR above the third quartile, it's an **extreme outlier**.
- The value 25 is also greater than 21.5, so it is a **mild outlier**.
- The value 19 is the **highest value that falls within the included range**.

So when we draw our box-and-whisker plot, the whiskers will only go out as far as 2 and 19 respectively. The points outside of that range are all outliers. Here is the plot:



For a step-by-step guide to making a box plot by hand, click here.

Making Box-and-Whisker Plots Using a Graphing Calculator

Graphing calculators make analyzing large lists of data easy. They have built-in algorithms for finding the median and the quartiles, and can be used to display box-and-whisker plots.

Example D

The ages of all the passengers traveling in a train carriage are shown below.

35, 42, 38, 57, 2, 24, 27, 36, 45, 60, 38, 40, 40, 44, 1, 44, 48, 84, 38, 20, 4, 2, 48, 58, 3, 20, 6, 40, 22, 26, 17, 18, 40, 51, 62, 31, 27, 48, 35, 27, 37, 58, 21

Use a graphing calculator to:

a) *obtain the 5 number summary for the data.*

b) *create a box-and-whisker plot.*

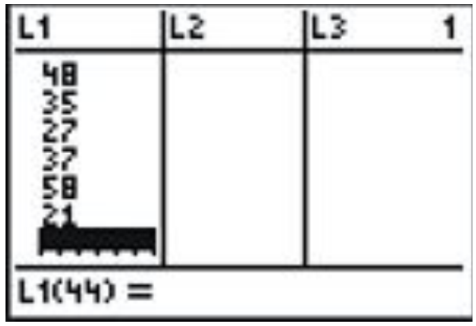
c) *determine if any of the points are outliers.*

Solution

Enter the data in your calculator:

Press **[START]** then choose **[EDIT]**.

Enter all 43 data points in list L_1.



Find the 5 number summary:

Press **[START]** again. Use the right arrow to choose **[CALU]**.

Highlight the **1-Var Stats** option. Press **[EDIT]**.

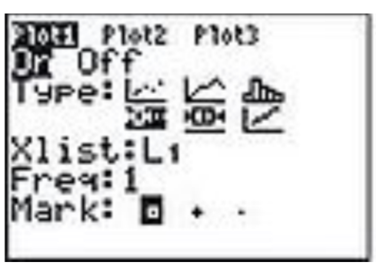
The single variable statistics summary appears.

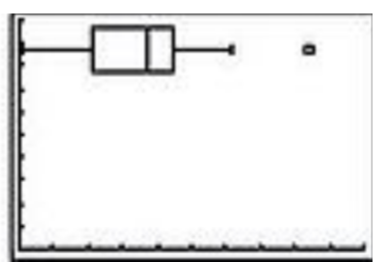
Note the **mean** ($\bar{x}$) is the first item given.

Use the down arrow to bring up the data for the five *number summary*. *n* is the number of data points, and the final fie numbers in the screen are the numbers we require.

TABLE 2.21:

	Symbol	Value
Lowest Value	**minX**	1
First Quartile	Q_2	21
Median	**Med**	37
Third Quartile	Q_3	45
Highest Value	**maxX**	84





Display the box-and-whisker plot:

Bring up the **[STARTPLOT]** option by pressing **[2nd]. [Y=]**.

Highlight **1:Plot1** and press **[ENTER]**.

There are two types of box-and-whisker plots available. The first automatically identifies outliers. Highlight it and press **[ENTER]**.

Press **[WINDOW]** and ensure that **Xmin** and **Xmax** allow for all data points to be shown. In this example, $Xmin = 0$ and $Xmax = 100$.

Press **[GRAPH]** and the box-and-whisker plot should appear.

The calculator will automatically identify outliers and plot them as such. You can use the **[TRACE]** function along with the arrows to identify outlier values. In this case there is one outlier: 84.

Watch this video for help with the Examples above.

Multimedia

MEDIA

Click image to the left for more content.

CK-12 Foundation: Box and Whisker Plots

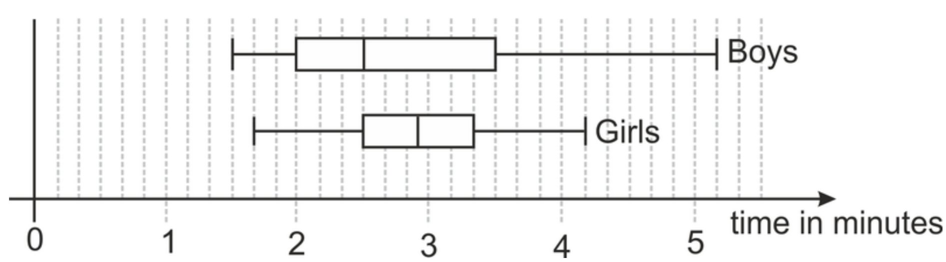
Vocabulary

- We call the division between the lower two quarters the **first quartile**. The division between the upper two quarters is the **third quartile** (the **second quartile** is, of course, the median).

- A **box-and-whisker plot** is formed by placing vertical lines at five positions, corresponding to the smallest value, the first quartile, the median, the third quartile and the greatest value. (These five numbers are often referred to as the **five number summary**.) A **box** is drawn between the position of the first and third quartiles, and horizontal line segments (the **whiskers**) connect the box with the two extreme values.

Guided Practice

The box-and-whisker plots below represent the times taken by a school class to complete an obstacle course. The times have been separated into boys and girls. The boys and the girls each think that they did best. Determine the five number summary for both the boys and the girls and give a convincing argument for each of them.



Solution

Comparing two sets of data with a box-and-whisker plot is relatively straightforward. For example, you can see that the data for the boys is more spread out, both in terms of the range and the inter-quartile range.

The five number summary for each is shown in the table below.

TABLE 2.22:

	Boys	Girls
Lowest value	1:30	1:40

TABLE 2.22: (continued)

	Boys	Girls
First Quartile	2:00	2:30
Median	2:30	2:55
Third Quartile	3:30	3:20
Highest value	5:10	4:10

Here are some points each side could use in their argument:

Boys:

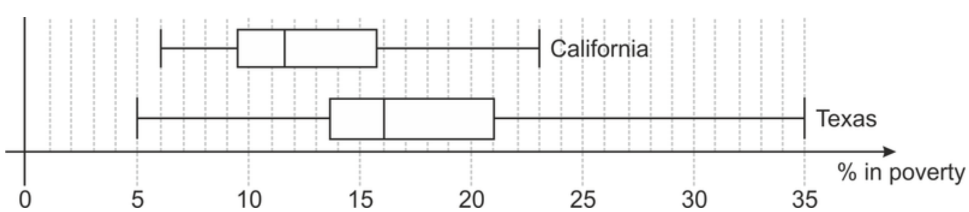
- The boys had the fastest time (1 minute 30 seconds), so the fastest individual was a boy.
- The boys also had the smaller median (2 minutes 30 seconds), meaning half of the boys were finished when only one fourth of the girls were finished (since the girls' first quartile is also 2:30). In other words, the boys' average time was faster.

Girls:

- The boys had the slowest time (5 minutes 10 seconds), so by the time all the girls were finished there was still at least one boy completing the course.
- The girls had the smaller third quartile (3 min 20 seconds), meaning that even without taking the slowest fourth of each group into account, the girls were still quickest.

Practice

1. Draw a box-and-whisker plot for the following unordered data: 49, 57, 53, 54, 57, 49, 67, 51, 57, 56, 59, 57, 50, 49, 52, 53, 50, 58
2. A simulation of a large number of runs of rolling 3 dice and adding the numbers results in the following 5-number summary: **3, 8, 10.5, 13, 18**. Make a box-and-whisker plot for the data and comment on the differences between it and the plot in example B.
3. The box-and-whisker plots below represent the percentage of people living below the poverty line by county in both Texas and California. Determine the 5-number summary for each state, and comment on the spread of each distribution.



4. The 5-number summary for the average daily temperature in Atlantic City, NJ^1 (given in $^\circ F$) is: **31, 39, 52, 68, 76**. Draw the box-and-whisker plot for this data and use it to determine which of the following, if any, would be considered outliers if they were included in the data:
   a. January's record high temperature of 78°
   b. January's record low temperature of -8°
   c. April's record high temperature of 94°
   d. The all time record high of 106°
5. In 1887 Albert Michelson and Edward Morley conducted an experiment to determine the speed of light. The data for the first 10 runs (5 results in each run) is given below. Each value represents how many kilometers per second over 299,000 km/s was measured. Create a box-and-whisker plot of the data. Be sure to identify

outliers and plot them as such. *850, 740, 900, 1070, 930, 850, 950, 980, 980, 880, 960, 940, 960, 940, 880, 800, 850, 880, 900, 840, 880, 880, 800, 860, 720, 720, 620, 860, 970, 950, 890, 810, 810, 820, 800, 770, 760, 740, 750, 760, 890, 840, 780, 810, 760, 810, 790, 810, 820, 850*

6. Is it possible to have outliers on both ends of a data set? Explain.
7. Is it possible for more than half the values in a data set to be outliers? Explain.
8. Is it possible for more than a quarter of the values in a data set to be outliers? Explain.
9. Is it possible for either of the whiskers in a box-and-whisker plot to be of zero length? Explain.
10. Is it possible for either of the whiskers in a box-and-whisker plot to be longer than the box? Explain.
11. Is it possible for either of the whiskers in a box-and-whisker plot to be twice as long as the box? Explain.

[1]Information taken from data published by Rutgers University Climate Lab (http://climate.rutgers.edu)

2.6 Double Box-and-Whisker Plots

Here you'll learn how to construct and interpret double box-and-whisker plots and use double box-and-whisker plots to solve problems. You'll also learn how to draw double box-and-whisker plots using Texas Instrument calculators.

A farmer is testing two new soil mixtures to see if his corn grows taller. He tests the two mixtures on similar patches of land with 50 plants each and compares the results. How can he organize the data to see which mixture will achieve the best results?

Watch This

First watch this video to learn about double box-and-whisker plots.

MEDIA

Click image to the left for more content.

CK-12 Foundation: Chapter8DoubleBoxandWhiskerPlotsA

Then watch this video to see some examples.

MEDIA

Click image to the left for more content.

CK-12 Foundation: Chapter8DoubleBoxandWhiskerPlotsB

Guidance

Double box-and-whisker plots give you a quick visual comparison of 2 sets of data, as was also found with other double graph forms covered in previous Concepts. The difference with double box-and-whisker plots is that you are also able to quickly visually compare the means, the medians, the maximums (upper range), and the minimums (lower range) of the data.

The double box-and-whisker plots in the first 2 examples were drawn using a program called Autograph. You can also draw double box-and-whisker plots by hand using pencil and paper or by using your TI-84 calculator, as was done in the third example.

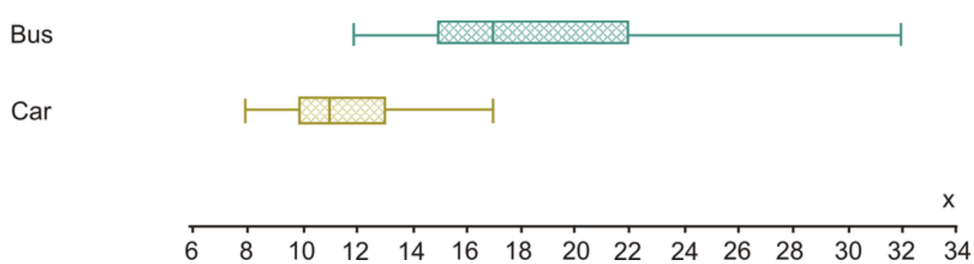
Example A

Emma and Daniel are surveying the times it takes students to arrive at school from home. There are 2 main groups of commuters who were in the survey. There were those who drove their own cars to school, and there were those who took the school bus. Emma and Daniel collected the following data:

Bus times (min)	14	18	16	22	25	12	32	16	15	18
Car times (min)	12	10	13	14	9	17	11	10	8	11

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to compare the times it takes for students to arrive at school either by car or by bus.

When plotted, the box-and-whisker plots look like the following:



Using the medians, 50% of the cars arrive at school in 11 minutes or less, whereas 50% of the students arrive by bus in 17 minutes or less. The range for the car times is $17-8=9$ minutes. For the bus times, the range is $32-12=20$ minutes. Since the range for the driving times is smaller, it means the times to arrive by car are less spread out. This would, therefore, mean that the times are more predictable and reliable.

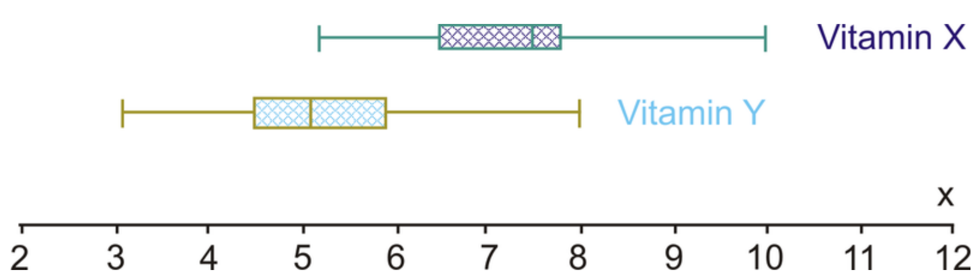
Example B

A new drug study was conducted by a drug company in Medical Town. In the study, 15 people were chosen at random to take Vitamin X for 2 months and then have their cholesterol levels checked. In addition, 15 different people were randomly chosen to take Vitamin Y for 2 months and then have their cholesterol levels checked. All 30 people had cholesterol levels between 8 and 10 before taking one of the vitamins. The drug company wanted to see which of the 2 vitamins had the greatest impact on lowering people's cholesterol. The following data was collected:

Vitamin X	7.2	7.5	5.2	6.5	7.7	10	6.4	7.6	7.7	7.8	8.1	8.3	7.2	7.1	6.5
Vitamin Y	4.8	4.4	4.5	5.1	6.5	8	3.1	4.6	5.2	6.1	5.5	4.2	4.5	5.9	5.2

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to compare the 2 vitamins and provide a conclusion for the drug company.

When plotted, the box-and-whisker plots look like the following:



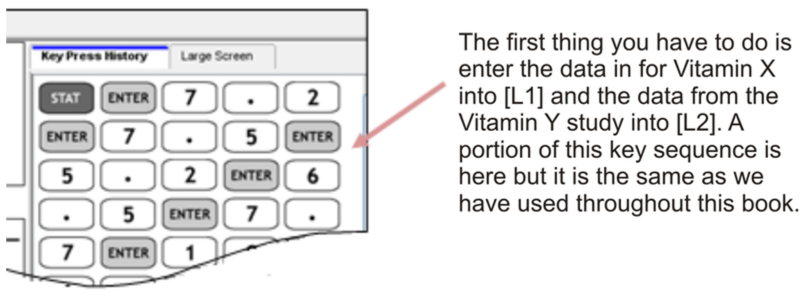
Using the medians, 50% of the people in the study had cholesterol levels of 7.5 or lower after being on Vitamin X for 2 months. Also 50% of the people in the study had cholesterol levels of 5.1 or lower after being on Vitamin Y for

2 months. Knowing that the participants of the survey had cholesterol levels between 8 and 10 before beginning the study, it appears that Vitamin Y had a bigger impact on lowering the cholesterol levels. The range for the cholesterol levels for people taking Vitamin X was $10 - 5.2 = 4.8$ points, while the range for the cholesterol levels for people taking Vitamin Y was $8 - 3.1 = 4.9$ points. Therefore, the range is not useful in making any conclusions.

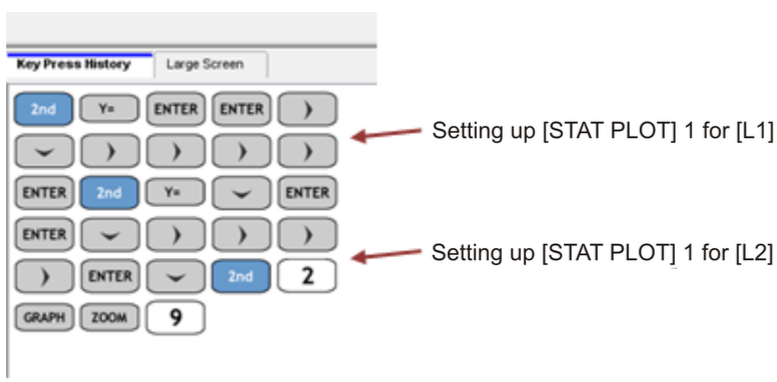
Example C

Draw the double box-and-whisker plots in Example B using the TI-84 calculator.

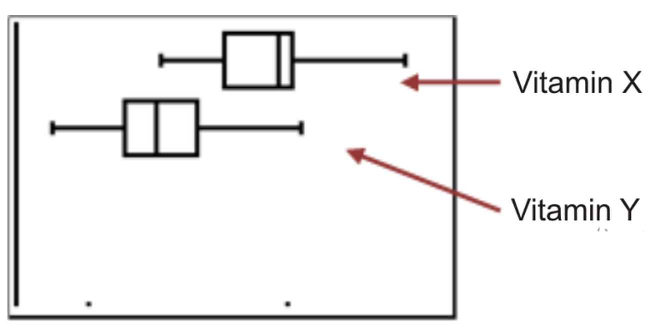
Follow the key sequence below to draw the double box-and-whisker plots:



After entering the data into L1 and L2, the next step is to graph the data by using STAT PLOT.



The resulting graph looks like the following:



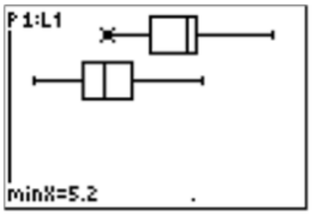
You can then press

TRACE

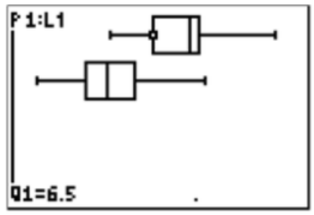
and find the five-number summary. The five-number summary is shown below for Vitamin X. By pressing the

▼

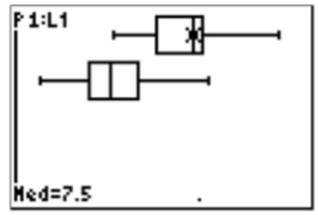
button, you can get to the second box-and-whisker plot (for Vitamin Y) and collect the five-number summary for this box-and-whisker plot.



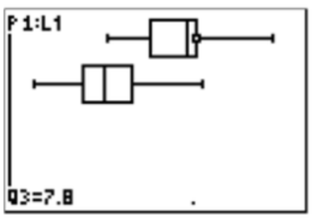
Minimum value



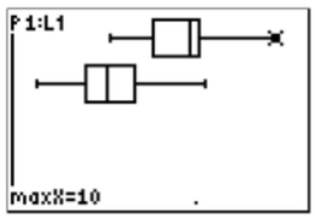
Lower Quartile value



Median value



Upper Quartile value



Maximum value

Points to Consider

- How is comparing double graphs (i.e., double box-and-whisker plots) useful when doing statistics?

Guided Practice

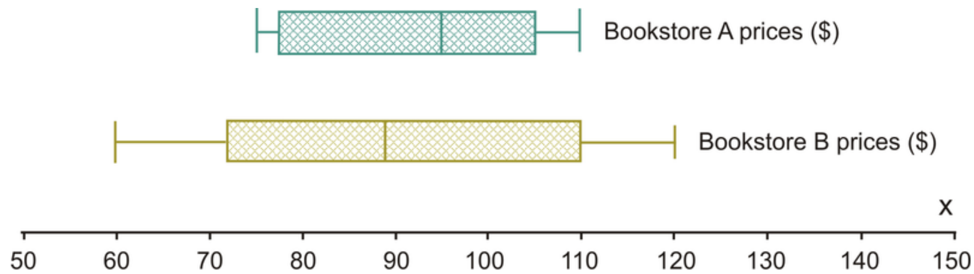
2 campus bookstores are having a price war on the prices of their first-year math books. James, a first-year math major, is going into each store to try to find the cheapest books he can find. He looks at 5 randomly chosen first-year books for first-year math courses in each store to determine where he should buy the 5 textbooks he needs for his courses this coming year. He collects the following data:

Bookstore A prices($)	95	75	110	100	80
Bookstore B prices($)	120	60	89	84	100

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to compare the 2 bookstores' prices, and provide a conclusion for James as to where to buy his books for his first-year math courses.

Answer:

The box-and-whisker plots are plotted and look like the following:



Using the medians, 50% of the books at Bookstore A are likely to be in the price range of $95 or less, whereas at Bookstore B, 50% of the books are likely to be around $89 or less. At first glance, you would probably recommend to James that he go to Bookstore B. Let's look at the ranges for the 2 bookstores to see the spread of data. For Bookstore A, the range is $\$110-\$75=\$35$, and for Bookstore B, the range is $\$120-\$60=\$60$. With the spread of the data much greater at Bookstore B than at Bookstore A, (i.e., the range for Bookstore B is greater than that for Bookstore A), to say that it would be cheaper to buy James's books at Bookstore A would be more predictable and reliable. You would, therefore, suggest to James that he is probably better off going to Bookstore A.

Interactive Practice

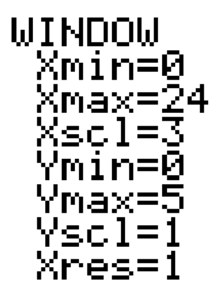
Practice

1. International Baccalaureate has 2 levels of courses, which are standard level (SL) and higher level (HL). Students say that study times are the same for both the standard level exams and the higher level exams. The following data represents the results of a survey conducted to determine how many hours a random sample of students studied for their final exams at each level:

HL Exams	15	16	16	17	19	10	5	6	5	5	8	10	8	12	17
SL Exams	10	6	6	7	9	12	2	6	2	5	7	20	18	8	18

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to determine the five-number summary for both sets of data. Compare the times students prepare for each level of exam.

2. Create the double box-and-whisker plots in question 1 using a TI calculator. Use the following WINDOW settings:

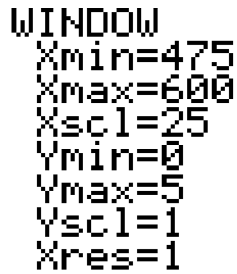
3. 

4. Students in the AP math class at BCU High School took their SATs for university entrance. The following scores were obtained for the math and verbal sections:

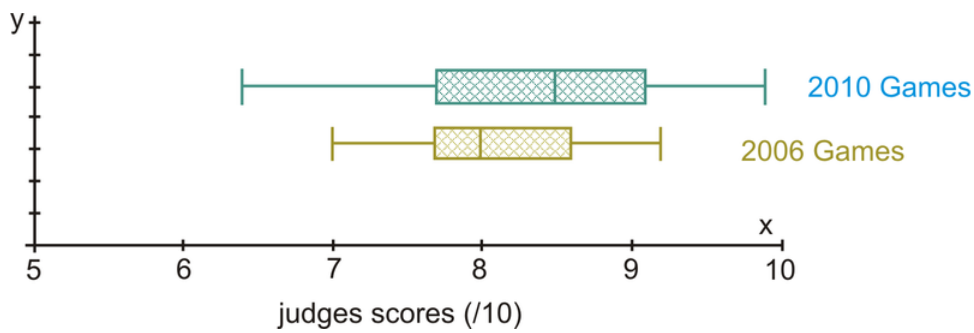
Math	529	533	544	562	513	519	560	575	568	537	561	522	563	571
Verbal	499	509	524	530	550	499	545	560	579	524	478	487	482	570

Draw a box-and-whisker plot for both sets of data on the same number line. Use the double box-and-whisker plots to determine the five-number summary for both sets of data. Compare the data for the 2 sections of the SAT using the five-number summary data.

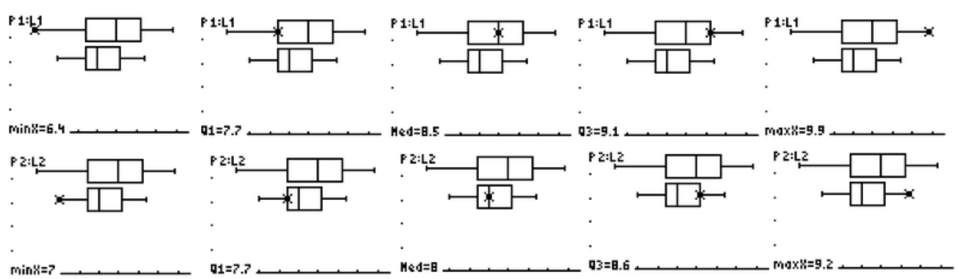
5. Create the double box-and-whisker plots in question 3 using a TI calculator. Use the following WINDOW settings:

6. 

The following box-and-whisker plots were drawn to analyze the data collected in a survey of scores for the doubles performances in the figure skating competitions at 2 Winter Olympic games. The box-and-whisker plot on the top represents the scores obtained at the 2010 winter games in Whistler, BC. The box-and-whisker plot on the bottom represents the scores obtained at the 2006 winter games in Torino, Italy.



Here is what the double box-and-whisker plots look like when created with a TI-84 calculator:



Use the double box-and-whisker plots to answer the following questions:

5. What is the five-number summary for the 2006 games?
6. What is the five-number summary for the 2010 games?
7. 50% of the judge's scores were at least what for the 2006 games? 50% of the judge's scores were at least what for the 2010 games?
8. What was the range for the scores in the 2006 games? What was the range for the scores in the 2010 games?
9. In which of the winter games would you say that the data is more equally spread out?
10. In which of the winter games would you say that the scoring was more predictable and reliable?

2.7 Summarizing Categorical Data

In previous lessons you worked with quantitative data involving variables such as rainfall and CEO salaries. In this lesson, you will analyze categorical data that involve variables such as movie preference and gender.

Guidance

Creating a Relative Frequency Table

The frequency table below shows the results of a survey that Olivia took at her school. She asked 50 randomly selected students whether they preferred action, comedy, or other types of movies.

TABLE 2.23:

Preferred Movie	Action	Comedy	Other	Total
Frequency	15	20	15	50

This information can be used to convert to a **relative frequency table** that uses decimals (see below):

TABLE 2.24:

Preferred Movie	Action	Comedy	Other	Total
Relative Frequency	$\frac{15}{50} = 0.3$	$\frac{20}{50} = 0.4$	$\frac{15}{50} = 0.3$	$\frac{50}{50} = 1.0$

The information can also be conveyed as percents (see below):

TABLE 2.25:

Preferred Movie	Action	Comedy	Other	Total
Relative Frequency	30%	40%	30%	100%

Notice that the Total column of a relative frequency table is always 1 or 100%.

In this example, the categorical variable is movie preference and the variable had three possible values: action, comedy, or other. The frequency table shown above has values for a single categorical variable. If you have two categorical variables, you list frequencies in a two-way table.

Creating a Two-Way Frequency Table

For her survey, Olivia also recorded the gender of each student. The results are shown in the two-way frequency table below. Each entry is the frequency of students who prefer a certain type of movie *and* are a certain gender.

TABLE 2.26:

	Action	Comedy	Other	Total
Girl	9	10	10	29
Boy	6	10	5	21
Total	**15**	**20**	**15**	**50**

Where have you seen the numbers in the total row before? Notice these are the numbers in our original relative frequency table without the gender variable.**frequency counts**. For example, 9 is the frequency count of girls that like action movies. The amount in the Total row Total column is the grand total of all values, or 50 in this specific survey.

Creating a Two-Way Relative Frequency Table

A **two-way relative frequency table** shows the joint relative frequencies and marginal relative frequencies.

A **joint relative frequency** is found by dividing the frequency count by the grand total. For example, $\frac{9}{50} = 0.18$ would be the joint relative frequency of students surveyed who are girls *and* preferred action movies. See the table below for all joint relative frequencies for Olivia's survey.

A **marginal relative frequency** is found by dividing a row total or column total by the grand total. See the table below **(in bold)** for the marginal relative frequencies for Olivia's survey.

TABLE 2.27:

	Action	Comedy	Other	**Total**
Girl	$\frac{9}{50} = 0.18$	$\frac{10}{50} = 0.2$	$\frac{10}{50} = 0.2$	$\frac{29}{50} = 0.58$
Boy	$\frac{6}{50} = 0.12$	$\frac{10}{50} = 0.2$	$\frac{5}{50} = 0.1$	$\frac{21}{50} = 0.42$
Total	$\frac{15}{50} = 0.3$	$\frac{20}{50} = 0.4$	$\frac{15}{50} = 0.3$	$\frac{50}{50} = 1.0$

What is the joint relative frequency of students surveyed who are girls and preferred comedies?

$\frac{10}{50} = 0.2$

What is the marginal relative frequency of students surveyed who are boys?

$\frac{21}{50} = 0.42$

Calculating Conditional Relative Frequencies

A **conditional relative frequency** is found by dividing a frequency count by the frequency's *row total or column total*. How does this differ from a joint relative frequency?

TABLE 2.28:

	Action	Comedy	Other	Total
Girl	9	10	10	29
Boy	6	10	5	21
Total	**15**	**20**	**15**	**50**

Find the conditional relative frequency that a student surveyed prefers action movies, given that the student is a boy. Divide the number of boys who preferred action movies by the number of boys. Round your answer to the nearest thousandth if necessary.

$\frac{6}{21} = 0.286$

28.6% of students preferred action movies, given that the student was a boy.

Find the conditional relative frequency that a student surveyed is a boy, given that the student prefers action movies. Divide the number of boys who prefer action movies by the number of students who preferred action movies. Round your answer to the nearest thousandth if necessary.

$\frac{6}{15} = 0.4$

40% of students surveyed were boys, given that the students preferred action movies.

You can also calculate conditional relative frequencies from a two way *relative* frequency tables. For example, to find the conditional relative frequency that a student surveyed is a boy, given that the student prefers action movies you can also use the joint and marginal relative frequencies: $\frac{0.12}{0.3} = 0.4$

Finding Possible Associations Between Variables

Olivia conducted her survey to because she was curious to see if gender would influence the type of movie student preferred. If gender had no influence, then the distribution of gender within each subgroup of movie preference would be roughly equal to the distribution of gender within the whole group.

First, we need to calculate the percent of all students surveyed who were girls.

58% of all students surveyed were girls. Where can you find this information on your two-way relative frequency table?

Second, we will take a look at the conditional relative frequencies.

Of the 15 students who preferred action movies, 9 are girls. The percent who are girls, given a preference for action movies is 60%.

Of the 20 students who preferred comedies, 10 are girls. The percent who are girls, given a preference for comedies is 50%.

Of the 15 students who preferred other types of movies, 10 are girls. The percent who are girls, given a preference for other types of movies is about 67%.

To analyze whether gender has an influence on the types of movies students prefer, we can compare the conditional relative frequencies to the percent of all students surveyed who were girls. If the percents are fairly close, gender would not appear to have an influence on movie choice.

The percent of girls among students who prefer action movies is fairly close to 58%, so gender does not appear to have an influence on preference for action movies.

The percent of girls among students who prefer comedies is lower than 58%, so gender does appear to have an influence on preference for comedies.

The percent of girls among students who prefer other types of movies is higher than 58%. What conclusion might you draw in this case?

What would happen if you analyzed gender influence by focusing on boys rather than girls?

For movie preference to be completely uninfluenced by gender, about how many girls would have to prefer each type of movie? Explain.

Guided Practice

Jack surveyed 30 of his classmates by asking them whether his or her favorite subject is math, English, or science. He also recorded the gender of each classmate surveyed. He recorded his results below.

TABLE 2.29:

	Math	English	Science	Total
Girl	6	6	6	18
Boy	4	4	4	12
Total	10	10	10	30

a. Create a two-way relative frequency table for the data using decimals.

TABLE 2.30:

	Math	English	Science	Total
Girl	0.2	0.2	0.2	0.6
Boy	0.133	0.133	0.133	0.4
Total	0.333	0.333	0.333	1.0

b. Find the joint relative frequency of surveyed students who are boys and prefer math.

$\frac{4}{30} = 0.133$

c. Find the marginal relative frequency of surveyed students who prefer English.

$\frac{10}{30} = 0.333$

d. Find the conditional relative frequency that a surveyed student prefers Science, given that the student is a girl.

$\frac{6}{10} = 0.6$

e. Does gender appear to influence subject preference in this case? Explain.

The percentage of students surveyed who were girls was 60%. Since all of the conditional relative frequencies for each subject are 60% (6/10), it would appear that gender does not have an influence on subject preference in this case.

Practice

Josh surveyed 80 of his classmates about their participation in school activities as well as whether they have a part-time job. The results are shown in the two-way frequency table below.

TABLE 2.31:

	Clubs only	Sports only	Both	Neither	Total
Part-time job	16	17	22	5	
No part-time job	4	6	6	4	
Total					

1. Complete the table by finding the row totals, column totals, and grand total.
2. Create a two-way relative frequency table using decimals.
3. Give the following relative frequencies as a percent:
   (a) The joint relative frequency of students surveyed who participate in sports only and have part-time jobs.
   (b) The marginal relative frequency of students surveyed who have a part time job.
   (c) The conditional relative frequency that a student surveyed partcipates in neither school clubs or sports, given that the student has a part-time job.
4. How do you think having a part-time job might influence participation in school activities? Support your position using data.

CHAPTER 3 Linear Equations and Inequalities

Chapter Outline

3.1 Two-Step Equations and Properties of Equality

Here you'll learn how to use the Addition Property and Multiplication Property to solve equations that you have written in two steps.

Suppose you are saving up to buy a leather jacket that costs \$250. You currently have \$90 put away, and you're saving at a rate of \$40 per month. Can you write an equation to represent this situation and then solve it in two steps to find the number of months needed to reach your goal? In this Concept, you'll learn how to solve problems just like this.

Guidance

Suppose Shaun weighs 146 pounds and wants to lose enough weight to wrestle in the 130-pound class. His nutritionist designed a diet for Shaun so he will lose about 2 pounds per week. How many weeks will it take Shaun to weigh enough to wrestle in his class?

This is an example that can be solved by working backward. In fact, you may have already found the answer by using this method. *The solution is 8 weeks*.

By translating this situation into an algebraic sentence, we can begin the process of **solving equations**. To solve an equation means to "undo" all the operations of the sentence, leaving a value for the variable.

Translate Shaun's situation into an equation.

$$-2w+146=130$$

This sentence has two operations: addition and multiplication. To find the value of the variable, we must use both properties of equality: the Addition Property of Equality and the Multiplication Property of Equality.

Procedure to Solve Equations of the Form $ax+b=$***some number:***

1. Use the Addition Property of Equality to get the variable term ax alone on one side of the equation:

$$ax = some\ number$$

2. Use the Multiplication Property of Equality to get the variable x alone on one side of the equation:

$$x = some\ number$$

Example A

Solve Shaun's problem.

Solution: $-2w+146=130$

Apply the Addition Property of Equality: $-2w+146-146=130-146$.

Simplify: $-2w=-16$.

Apply the Multiplication Property of Equality: $-2w \div -2 = -16 \div -2$.

The solution is $w = 8$.

It will take 8 weeks for Shaun to weigh 130 pounds.

Solving Equations by Combining Like Terms

Michigan has a 6% sales tax. Suppose you made a purchase and paid \$95.12, including tax. How much was the purchase before tax?

Begin by determining the noun that is unknown and choose a letter as its representation.

The purchase price is unknown so this is our variable. Call it p. Now translate the sentence into an algebraic equation.

$$price + (0.06)price = total\ amount$$
$$p + 0.06p = 95.12$$

To solve this equation, you must know how to **combine like terms.**

Like terms are expressions that have **identical** variable parts.

According to this definition, you can only combine like terms if they are identical. **Combining like terms only applies to addition and subtraction!** This is not a true statement when referring to multiplication and division.

The numerical part of an algebraic term is called the **coefficient**. To combine like terms, you add (or subtract) the coefficients of the identical variable parts.

Example B

Identify the like terms, and then combine.

$$10b + 7bc + 4c + (-8b)$$

Solution: Like terms have identical variable parts. The only terms having identical variable parts are $10b$ and $-8b$. To combine these like terms, add them together.

$$10b + 7bc + 4c + -8b = 2b + 7bc + 4c$$

You will now apply this concept to the Michigan sales tax situation.

Example C

What was the purchase amount from this section's opening scenario?

Solution: $p + 0.06p = 95.12$

Combine the like terms: $p + 0.06p = 1.06p$, since $p = 1p$.

Simplify: $1.06p = 95.12$.

Apply the Multiplication Property of Equality: $1.06p \div 1.06 = 95.12 \div 1.06$.

Simplify: $p = 89.74$.

The price before tax was \$89.74.

Example D

Sometimes you will see equations with coefficients as letters.

Solve for x: $ax+7=12$

$ax+7=12$

$ax=5$

$x=\frac{5}{a}$

The next several examples show how algebraic equations can be created to represent real-world situations.

Video Review

MEDIA

Click image to the left for more content.

MEDIA

Click image to the left for more content.

Guided Practice

1. *An emergency plumber charges* \$65 *as a call-out fee plus an additional* \$75 per hour. He arrives at a house at 9:30 and works to repair a water tank. If the total repair bill is \$196.25, at what time was the repair completed?

2. *To determine the temperature in Fahrenheit, multiply the Celsius temperature by 1.8 and then add 32. Determine the Celsius temperature if it is* $89^\circ F$.

Solutions:

1. Translate the sentence into an equation. The number of hours it took to complete the job is unknown, so call it h.

Write the equation: $65+75(h)=196.25$.

Apply the Addition Property and simplify.

$$65+75(h)-65=196.25-65$$

$$75(h) = 131.25$$

Apply the Multiplication Property of Equality: $75(h) \div 75 = 131.25 \div 75$.

Simplify: $h = 1.75$.

The plumber worked for 1.75 hours, or 1 hour, 45 minutes. Since he started at 9:30, the repair was completed at 11:15.

2. Translate the sentence into an equation. The temperature in Celsius is unknown; call it C.

Write the equation: $1.8C + 32 = 89$.

Apply the Addition Property and simplify.

$$1.8C + 32 - 32 = 89 - 32$$
$$1.8C = 57$$

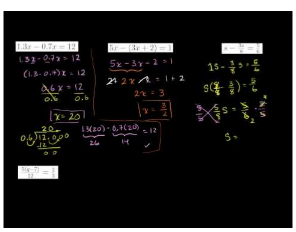
Apply the Multiplication Property of Equality: $1.8C \div 1.8 = 57 \div 1.8$.

Simplify: $C = 31.67$.

If the temperature is $89^\circ F$, then it is $31.67^\circ C$.

Practice

Sample explanations for some of the practice exercises below are available by viewing the following video. Note that there is not always a match between the number of the practice exercise in the video and the number of the practice exercise listed in the following exercise set. However, the practice exercise is the same in both. CK-12 Basic Algebra: Two-Step Equations (13:50)



MEDIA

Click image to the left for more content.

1. Define *like terms*. Give an example of a pair of like terms and a pair of unlike terms.
2. Define *coefficient*.

In 3 – 7, combine the like terms.

3. $-7x + 39x$
4. $3x^2 + 21x + 5x + 10x^2$
5. $6xy + 7y + 5x + 9xy$
6. $10ab + 9 - 2ab$
7. $-7mn - 2mn^2 - 2mn + 8$
8. Explain the procedure used to solve $-5y - 9 = 74$.

Solve and check your solution.

9. $1.3x - 0.7x = 12$
10. $6x - 1.3 = 3.2$
11. $5x - (3x + 2) = 1$
12. $4(x + 3) = 1$
13. $5q - 7 = \frac{2}{3}$
14. $\frac{3}{5}x + \frac{5}{2} = \frac{2}{3}$
15. $s - \frac{3s}{8} = \frac{5}{6}$
16. $0.1y + 11 = 0$
17. $\frac{5q-7}{12} = \frac{2}{3}$
18. $\frac{5(q-7)}{12} = \frac{2}{3}$
19. $33t - 99 = 0$
20. $5p - 2 = 32$
21. $14x + 9x = 161$
22. $3m - 1 + 4m = 5$
23. $8x + 3 = 11$
24. $24 = 2x + 6$
25. $66 = \frac{2}{3}k$
26. $\frac{5}{8} = \frac{1}{2}(a + 2)$
27. $16 = -3d - 5$
28. Solve for x: $ax - 8 = -2$
29. Solve for y: $-5 + ay = 12$
30. Solve for x: $-ax - \frac{1}{2} = \frac{3}{4}$
31. Jayden purchased a new pair of shoes. Including a 7% sales tax, he paid $84.68. How much did his shoes cost before sales tax?
32. A mechanic charges $98 for parts and $60 per hour for labor. Your bill totals $498.00, including parts and labor. How many hours did the mechanic work?
33. An electric guitar and amp set costs $1195.00. You are going to pay $250 as a down payment and pay the rest in 5 equal installments. How much should you pay each month?
34. Jade is stranded downtown with only $10 to get home. Taxis cost $0.75 per mile, but there is an additional $2.35 hire charge. Write a formula and use it to calculate how many miles she can travel with her money. Determine how many miles she can ride.
35. Jasmin's dad is planning a surprise birthday party for her. He will hire a bouncy castle and provide party food for all the guests. The bouncy castle costs $150 dollars for the afternoon, and the food will cost $3.00 per person. Andrew, Jasmin's dad, has a budget of $300. Write an equation to help him determine the maximum number of guests he can invite.

3.2 Equations with Variables on Both Sides

Here you'll learn how to manipulate equations with variables on both sides of the equal sign so that all variable terms appear on one side.

What if you had an equation like $9(x-2) = 6-3x$ in which the variable was on both sides of the equal sign? How could you solve this equation for x? After completing this Concept, you'll be able to solve equations like this one where the variable is on both sides.

Watch This

MEDIA

Click image to the left for more content.

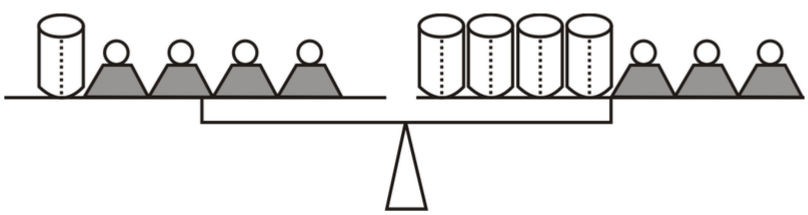
CK-12 Foundation: 0309S Equations with Variables on Both Sides

Guidance

When a variable appears on both sides of the equation, we need to manipulate the equation so that all variable terms appear on one side, and only constants are left on the other.

Example A

Dwayne was told by his chemistry teacher to measure the weight of an empty beaker using a balance. Dwayne found only one lb weights, and so devised the following way of balancing the scales.



Knowing that each weight is one lb, calculate the weight of one beaker.

Solution

We know that the system balances, so the weights on each side must be equal. We can write an algebraic expression based on this fact. The unknown quantity, the weight of the beaker, will be our x. We can see that on the left hand scale we have one beaker and four weights. On the right scale, we have four beakers and three weights. The balancing of the scales is analogous to the balancing of the following equation:

$$x+4 = 4x+3$$

"One beaker plus 4 lbs **equals** 4 beakers plus 3 lbs"

To solve for the weight of the beaker, we want all the constants (numbers) on one side and all the variables (terms with x in them) on the other side. Since there are more beakers on the right and more weights on the left, we'll try to move all the x terms (beakers) to the right, and the constants (weights) to the left.

First we subtract 3 from both sides to get $x+1=4x$.

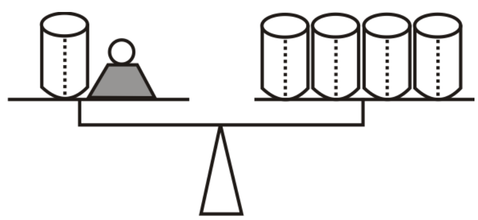
Then we subtract x from both sides to get $1=3x$.

Finally we divide by 3 to get $\frac{1}{3}=x$.

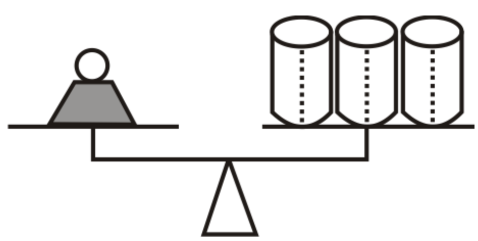
The weight of the beaker is one-third of a pound.

We can do the same with the real objects as we did with the equation. Just as we subtracted amounts from each side of the equation, we could remove a certain number of weights or beakers from each scale. Because we remove the same number of objects from each side, we know the scales will still balance.

First, we could remove three weights from each scale. This would leave one beaker and one weight on the left and four beakers on the right (in other words $x+1=4x$):



Then we could remove one beaker from each scale, leaving only one weight on the left and three beakers on the right, to get $1=3x$:



Looking at the balance, it is clear that the weight of one beaker is one-third of a pound.

To see more examples of solving equations with variables on both sides of the equation, see the Khan Academy video at http://www.youtube.com/watch?v=Zn-GbH2S0Dk.

Solve an Equation with Grouping Symbols

As you've seen, we can solve equations with variables on both sides even when some of the variables are in parentheses; we just have to get rid of the parentheses, and then we can start combining like terms. We use the same technique when dealing with fractions: first we multiply to get rid of the fractions, and then we can shuffle the terms around by adding and subtracting.

Example B

Solve $3x+2=\frac{5x}{3}$.

Solution

The first thing we'll do is get rid of the fraction. We can do this by multiplying both sides by 3, leaving $3(3x+2)=5x$.

Then we distribute to get rid of the parentheses, leaving $9x+6=5x$.

We've already got all the constants on the left side, so we'll move the variables to the right side by subtracting $9x$ from both sides. That leaves us with $6=-4x$.

And finally, we divide by -4 to get $-\frac{3}{2}=x$, or $x=-1.5$.

Example C

Solve the following equation for x: $\frac{14x}{(x+3)}=7$

Solution

The form of the left hand side of this equation is known as a **rational function** because it is the ratio of two other functions: $14x$ and $(x+3)$. But we can solve it just like any other equation involving fractions.

First we multiply both sides by $(x+3)$ to get rid of the fraction. Now our equation is $14x=7(x+3)$.

Then we distribute: $14x=7x+21$.

Then subtract $7x$ from both sides: $7x=21$.

And divide by 7: $x=3$.

Solve Real-World Problems Using Equations with Variables on Both Sides

Here's another chance to practice translating problems from words to equations. What is the equation asking? What is the **unknown** variable? What quantity will we use for our variable?

The text explains what's happening. Break it down into small, manageable chunks, and follow what's going on with our variable all the way through the problem.

More on Ohm's Law

Recall that the electrical current, I (amps), passing through an electronic component varies directly with the applied voltage, V (volts), according to the relationship $V=I\cdot R$ where R is the resistance measured in Ohms (Ω).

The resistance R of a number of components wired in a **series** (one after the other) is simply the sum of all the resistances of the individual components.

Example D

In an attempt to find the resistance of a new component, a scientist tests it in series with standard resistors. A fixed voltage causes a 4.8 amp current in a circuit made up from the new component plus a 15Ω *resistor in series. When the component is placed in a series circuit with a* 50Ω *resistor, the same voltage causes a 2.0 amp current to flow. Calculate the resistance of the new component.*

This is a complex problem to translate, but once we convert the information into equations it's relatively straightforward to solve. First, we are trying to find the resistance of the new component (in Ohms, Ω). This is our x. We don't know the voltage that is being used, but we can leave that as a variable, V. Our first situation has a total resistance that equals the unknown resistance plus 15Ω. The current is 4.8 amps. Substituting into the formula $V=I\cdot R$, we get $V=4.8(x+15)$.

Our second situation has a total resistance that equals the unknown resistance plus 50Ω. The current is 2.0 amps. Substituting into the same equation, this time we get $V=2(x+50)$.

We know the voltage is fixed, so the V in the first equation must equal the V in the second. That means we can set the right-hand sides of the two equations equal to each other: $4.8(x+15)=2(x+50)$. Then we can solve for x.

Distribute the constants first: $4.8x+72=2x+100$.

Subtract $2x$ from both sides: $2.8x+72 = 100$.

Subtract 72 from both sides: $2.8x = 28$.

Divide by 2.8: $x = 10$.

The resistance of the component is 10Ω.

Watch this video for help with the Examples above.

MEDIA

Click image to the left for more content.

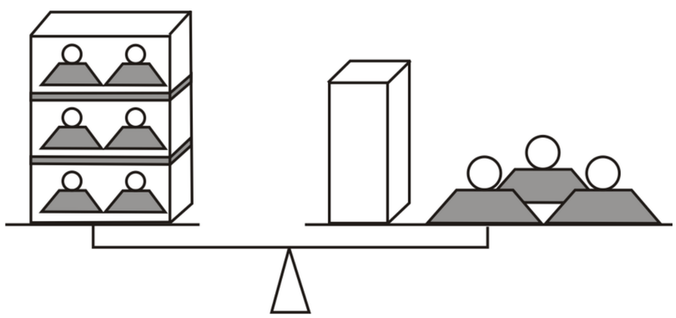
CK-12 Foundation: Variables on Both Sides

Vocabulary

An equation has the **variable on both sides** if the variable appears somewhere on each side of the equation. Distribute as necessary and then simplify the equation to have the unknown on only one side.

Guided Practice

1. *Sven was told to find the weight of an empty box with a balance. Sven found some one lb weights and five lb weights. He placed two one lb weights in three of the boxes and with a fourth empty box found the following way of balancing the scales:*



Knowing that small weights are one lb and big weights are five lbs, calculate the weight of one box.

2. *Solve* $7x+2 = \frac{5x-3}{6}$.

Solutions:

1. We know that the system balances, so the weights on each side must be equal. We can write an algebraic expression based on this equality. The unknown quantity—the weight of each empty box, in pounds—will be our x. A box with two 1 lb weights in it weighs $(x+2)$ pounds. Our equation, based on the picture, is $3(x+2) = x+3(5)$.

Distributing the 3 and simplifying, we get $3x+6 = x+15$.

Subtracting x from both sides, we get $2x+6 = 15$.

Subtracting 6 from both sides, we get $2x = 9$.

And finally we can divide by 2 to get $x = \frac{9}{2}$, or $x = 4.5$.

Each box weighs 4.5 lbs.

2. Again we start by eliminating the fraction. Multiplying both sides by 6 gives us $6(7x+2) = 5x-3$, and distributing gives us $42x+12 = 5x-3$.

Subtracting $5x$ from both sides gives us $37x+12 = -3$.

Subtracting 12 from both sides gives us $37x = -15$.

Finally, dividing by 37 gives us $x = -\frac{15}{37}$.

Practice

For 1-11, solve the following equations for the unknown variable.

1. $3(x-1) = 2(x+3)$
2. $7(x+20) = x+5$
3. $9(x-2) = 3x+3$
4. $2\left(a-\frac{1}{3}\right) = \frac{2}{5}\left(a+\frac{2}{3}\right)$
5. $\frac{2}{7}\left(t+\frac{2}{3}\right) = \frac{1}{5}\left(t-\frac{2}{3}\right)$
6. $\frac{1}{7}\left(v+\frac{1}{4}\right) = 2\left(\frac{3v}{2}-\frac{5}{2}\right)$
7. $\frac{y-4}{11} = \frac{2}{5} \cdot \frac{2y+1}{3}$
8. $\frac{z}{16} = \frac{2(3z+1)}{9}$
9. $\frac{q}{16} + \frac{q}{6} = \frac{(3q+1)}{9} + \frac{3}{2}$
10. $\frac{3}{x} = \frac{2}{x+1}$
11. $\frac{5}{2+p} = \frac{3}{p-8}$
12. Manoj and Tamar are arguing about a number trick they heard. Tamar tells Andrew to think of a number, multiply it by five and subtract three from the result. Then Manoj tells Andrew to think of a number, add five and multiply the result by three. Andrew says that whichever way he does the trick he gets the same answer.
    a. What was the number Andrew started with?
    b. What was the result Andrew got both times?
    c. Name another set of steps that would have resulted in the same answer if Andrew started with the same number.
13. Manoj and Tamar try to come up with a harder trick. Manoj tells Andrew to think of a number, double it, add six, and then divide the result by two. Tamar tells Andrew to think of a number, add five, triple the result, subtract six, and then divide the result by three.
    a. Andrew tries the trick both ways and gets an answer of 10 each time. What number did he start out with?
    b. He tries again and gets 2 both times. What number did he start out with?
    c. Is there a number Andrew can start with that will *not* give him the same answer both ways?
    d. **Bonus:** Name another set of steps that would give Andrew the same answer every time as he would get from Manoj's and Tamar's steps.
14. I have enough money to buy five regular priced CDs and have \$6 left over. However, all CDs are on sale today, for \$4 less than usual. If I borrow \$2, I can afford nine of them.
    a. How much are CDs on sale for today?
    b. How much would I have to borrow to afford nine of them if they weren't on sale?
15. Five identical electronics components were connected in series. A fixed but unknown voltage placed across them caused a 2.3 amp current to flow. When two of the components were replaced with standard 10Ω resistors, the current dropped to 1.9 amps. What is the resistance of each component?
16. Solve the following resistance problems. Assume the same voltage is applied to all circuits.
    a. Three unknown resistors plus 20Ω give the same current as one unknown resistor plus 70Ω.

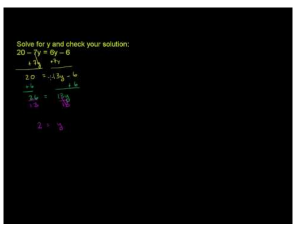
b. One unknown resistor gives a current of 1.5 amps and a 15Ω resistor gives a current of 3.0 amps.
c. Seven unknown resistors plus 18Ω gives twice the current of two unknown resistors plus 150Ω.
d. Three unknown resistors plus 1.5Ω gives a current of 3.6 amps and seven unknown resistors plus seven 12Ω resistors gives a current of 0.2 amps.

3.3 Solving Multi-Step Equations

Here, you will solve multi-step linear equations involving the Distributive Property, combining like terms or having the variable on both sides of the equals sign.

Let's look at cell phone companies again. Cell-U-Lar includes 500 texts in their \$79/month fee and then charges \$0.10 for each additional text. We-B-Mobile also includes 500 texts in their \$59/month fee and then charges \$0.25 for each additional text. At how many texts, over 500, will the two bills be the same?

Watch This



MEDIA

Click image to the left for more content.

Khan Academy: Multi-step equations 1

Guidance

The types of equations in this concept involve at least three steps. Keep in mind that the last two steps when solving a linear equation will always be the same: add or subtract the number that is on the same side of the equals sign as the variable, then multiply or divide by the number with the variable.

Example A

Solve $3(x-5)+4=10$.

Solution: When solving more complicated equations, start with one side and simplify as much as you can. The left side of this equation looks more complicated, so let's simplify it by using the Distributive Property and combining like terms.

$$\begin{aligned} 3(x-5)+4&=10\\ 3x-15+4&=10\\ 3x-11&=10 \end{aligned}$$

Now, this looks like an equation from the previous concept. Continue to solve.

$$\begin{aligned} 3x - \cancel{11} &= 10 \\ +\cancel{11} &\; +11 \\ \hline \frac{\cancel{3}x}{\cancel{3}} &= \frac{21}{3} \\ x &= 7 \end{aligned}$$

Check your answer: $3(7-5)+4 = 3 \cdot 2 + 4 = 6 + 4 = 10$

You can use properties to explain each step when solving an equation. There is usually more than one way to do so. For example, you can solve the equation in Example A both with and without the distributive property. The substitution property can be used where you combine like terms.

$3(x-5)+4 = 10$

$3x - 15 + 4 = 10$ Distributive Property

$3x - 11 = 10$ Substitution (think of this as substituting -11 in for -15+4)

$3x = 21$ Addition Property of Equality

$x = 7$ Division Property of Equality

Here is another way to explain your reasoning when solving an equation:

$3(x-5)+4 = 10$

$3(x-5) = 6$ Subtraction Property of Equality

$(x-5) = 2$ Division Property of Equality

$x = 7$ Addition Property of Equality

Example B

Solve $8x - 17 = 4x + 23$.

Solution: This equation has x on both sides of the equals sign. Therefore, we need to move one of the x terms to the other side of the equation. It does not matter which x term you move. We will move the $4x$ to the other side so that, when combined, the x term is positive.

$$\begin{aligned} 8x - 17 &= \cancel{4x} + 23 \\ -4x \quad &\; - \cancel{4x} \\ \hline 4x - \cancel{17} &= 23 \\ +\cancel{17} &\; +17 \\ \hline \frac{\cancel{4}x}{\cancel{4}} &= \frac{40}{4} \\ x &= 10 \end{aligned}$$

Check your answer:

$$\begin{aligned} 8(10) - 17 &= 4(10) + 23 \\ 80 - 17 &= 40 + 23 \\ 63 &= 63 \end{aligned}$$

Example C

Solve $2(3x-1)+2x=5-(2x-3)$.

Solution: This equation combines the two previous examples. First, use the Distributive Property.

$$\begin{aligned} 2(3x-1)+2x &= 5-(2x-3) \\ 6x-2+2x &= 5-2x+3 \end{aligned}$$

Don't forget to distribute the negative sign in front of the second set of parenthesis. Treat it like distributing a –1. Now, combine like terms and solve the equation.

$$\begin{aligned} 8x-\cancel{2} &= 8-\cancel{2x} \\ \underline{+2x+\cancel{2}} &\underline{+2+\cancel{2x}} \\ \frac{\cancel{10}x}{\cancel{10}} &= \frac{10}{10} \\ x &= 1 \end{aligned}$$

Check your answer:

$$\begin{aligned} 2(3(1)-1)+2(1) &= 5-(2(1)-3) \\ 2\cdot 2+2 &= 5-(-1) \\ 4+2 &= 6 \end{aligned}$$

Intro Problem Revisit Let's write an expression for the total cost associated with each cell phone company.

Cell-U-Lar: $79+0.1t$

We-B-Mobile: $59+0.25t$

Where t is the number of texts over 500. Set the two expressions equal to each other to see when the plans are equal.

$$\begin{aligned} 79+0.1t &= 59+0.25t \\ 20 &= 0.15t \\ 133.\bar{3} &= t \end{aligned}$$

At 133 additional texts, the bills would be just about the same, but don't forget that each company includes 500 texts at no additional cost. So, you would have to go just over 633 texts to make the two plans equal. If you text less than that, the plan from We-B-Mobile is a better deal. More than 633, then Cell-U-Lar makes more sense.

Guided Practice

Solve each equation below and check your answer.

1. $\frac{3}{4}+\frac{2}{3}x=2x+\frac{5}{6}$
2. $0.6(2x-7)=5x-5.1$

Answers

1. Use the LCD method introduced in the problem set from the previous concept. Multiply every term by the LCD of 4, 3, and 6.

$$\begin{aligned} 12\left(\frac{3}{4}+\frac{2}{3}x=2x+\frac{5}{6}\right) \\ 9+8x=24x+10 \end{aligned}$$

Now, combine like terms. Follow the steps from Example B.

$$\begin{aligned} 9+\cancel{8x}&=24x+\cancel{10} \\ \underline{-10-\cancel{8x}}&\underline{\;\;-8x-\cancel{10}} \\ \frac{-1}{16}&=\frac{\cancel{16}x}{\cancel{16}} \\ -\frac{1}{16}&=x \end{aligned}$$

Check your answer:

$$\begin{aligned} \frac{3}{4}+\frac{2}{3}\left(-\frac{1}{16}\right)&=2\left(-\frac{1}{16}\right)+\frac{5}{6} \\ \frac{3}{4}-\frac{1}{24}&=-\frac{1}{8}+\frac{5}{6} \\ \frac{18}{24}-\frac{1}{24}&=-\frac{3}{24}+\frac{20}{24} \\ \frac{17}{24}&=\frac{17}{24} \end{aligned}$$

2. Even though there are decimals in this problem, we can approach it like any other problem. Use the Distributive Property and combine like terms.

$$\begin{aligned} 0.6(2x+7)&=4.3x-5.1 \\ 1.2x+\cancel{4.2}&=\cancel{4.3x}-5.1 \\ \underline{-4.3x-\cancel{4.2}}&\underline{\;\;-\cancel{4.3x}-4.2} \\ \frac{\cancel{-3.1}x}{\cancel{-3.1}}&=\frac{-9.3}{-3.1} \\ x&=3 \end{aligned}$$

Check your answer:

$$\begin{aligned} 0.6(2(3)+7)&=4.3(3)-5.1 \\ 0.6\cdot 13&=12.9-5.1 \\ 7.8&=7.8 \end{aligned}$$

Practice

Solve each equation and check your solution. For numbers 1, 3 & 9 use mathematical properties to justify each step in the process.

1. $-6(2x-5)=18$
2. $2-(3x+7)=-x+19$
3. $3(x-4)=5(x+6)$
4. $x-\frac{4}{5}=\frac{2}{3}x+\frac{8}{15}$
5. $8-9x-5=x+23$
6. $x-12+3x=-2(x+18)$
7. $\frac{5}{2}x+\frac{1}{4}=\frac{3}{4}x+2$
8. $5\left(\frac{x}{3}+2\right)=\frac{32}{3}$
9. $7(x-20)=x+4$
10. $\frac{2}{7}\left(x+\frac{2}{3}\right)=\frac{1}{5}\left(2x-\frac{2}{3}\right)$
11. $\frac{1}{6}(x+2)=2\left(\frac{3x}{2}-\frac{5}{4}\right)$
12. $-3(-2x+7)-5x=8(x-3)+17$

Challenge Solve the equations below. Check your solution.

13. $\frac{3x}{16}+\frac{x}{8}=\frac{3x+1}{4}+\frac{3}{2}$
14. $\frac{x-1}{11}=\frac{2}{5}\cdot\frac{x+1}{3}$
15. $\frac{3}{x}=\frac{2}{x+1}$

3.4 Solving Equations with Rational Numbers

Introduction

The New Piece of Music

Jose has been playing the French horn for many years and until now, everything has been easy. Now Mrs. Kline, the band director has assigned him a new piece of music to work on and it is very tricky. Jose has been practicing the new piece.

His best rehearsal was on Saturday, when he practiced for 90 minutes. On Sunday, he had a birthday party to go to, so he did not practice as long. On Monday, he had a math test to study for and so he practiced for half as long as he did on Monday.

When Jose went to band practice on Tuesday afternoon, he struggled through the piece.

"How long did you practice?" Mrs. Kline asked him.

"Well, from Saturday to Tuesday I practiced a total of 3 hours," Jose said.

If this is true, how long did Jose practice on Monday and Tuesday? You will need to write an equation and solve it to figure out the answer to this problem. Jose needs to practice his French horn a bit more and you will need to use the information taught in this lesson to help you figure out each dilemma.

What You Will Learn

In this lesson, you will learn to perform the following skills.

- Solve multi-step equations involving decimals.
- Solve multi-step equations involving fractions.
- Solve multi-step equations involving multiple forms of rational numbers.
- Model and solve real-world problems using multi-step equations involving rational numbers.

Teaching Time

I. <u>**Solve Multi-Step Equations Involving Decimals**</u>

So far, we have focused on solving multi-step equations that contain integers. ***Integers*** **include positive whole numbers (1, 2, 3, 4, 5, . . .), their opposites (-1, -2, -3, -4, -5, . . .), and zero.**

Integers are examples of ***rational numbers***. **A rational number is any number that can be written as the ratio of two integers or you can think of this in fraction form.** So, an integer such as -3, which can be written as the ratio $\frac{-3}{1}$, is a rational number.

What are some other examples of rational numbers?

A ***fraction,*** such as $\frac{1}{4}$, can obviously be written as the ratio of two integers. So, fractions are rational numbers.

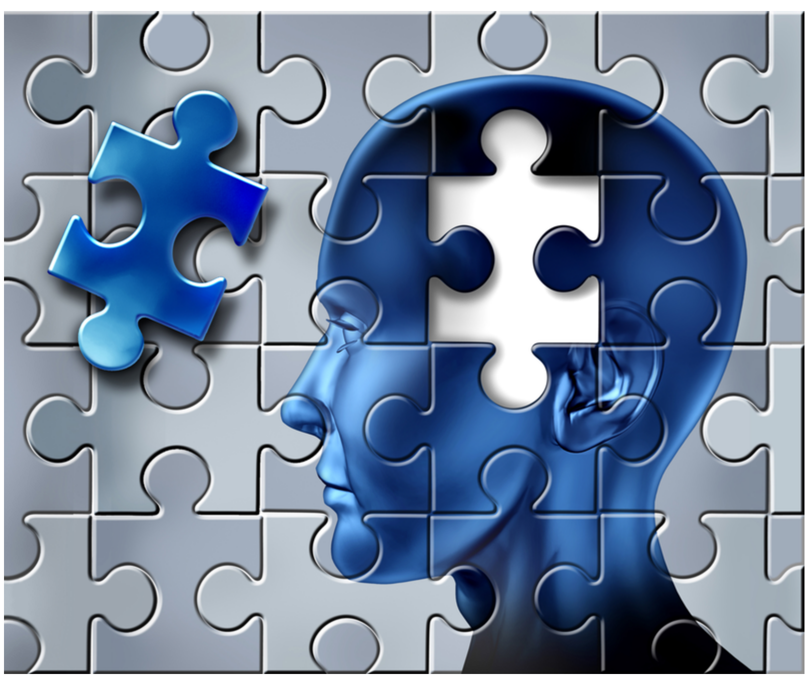
A terminating decimal, such as 0.1, is also rational because it can be written as the ratio $\frac{1}{10}$. A ***repeating decimal,*** such as $0.\overline{3}$, is rational because even though the digit 3 repeats over and over in the decimal form, it can be expressed as the ratio of two integers: $\frac{1}{3}$.

All integers, fractions, terminating decimals and repeating decimals are rational numbers.

Just as you have been solving equations with integers in them, we can also solve equations with other rational numbers in them.

Let's start by looking at solving equations involving decimals.

You will use the same strategy to solve a multi-step equation that includes decimals that you would to solve a multi-step equation that includes only integers. You will first combine like terms or use the distributive property to simplify the equation. Then, you will use inverse operations to isolate the variable on one side of the equation.



Remember back, you will need to remember how to perform operations involving decimals to be effective at solving equations with decimals.

Example

Solve for x: $3x - 2.5x + 0.5 = 4.5$

First, subtract the like terms—$3x$ and $2.5x$—on the left side of the equation. It may help to remember that $3x = 3.0x$.

$$\begin{aligned}3x - 2.5x + 0.5 &= 4.5\\(3.0x - 2.5x) + 0.5 &= 4.5\\0.5x + 0.5 &= 4.5\end{aligned}$$

Notice that 0.5 cannot be combined with $0.5x$ because they are not like terms.

Now, we can solve as we would solve any two-step equation.

The next step is to isolate the term with the variable, $0.5x$, on one side of the equation. Since 0.5 is *added* to $0.5x$, we should subtract 0.5 from both sides of the equation.

$$\begin{gathered}0.5x + 0.5 = 4.5\\0.5x + 0.5 - 0.5 = 4.5 - 0.5\\0.5x + 0 = 4.0\\0.5x = 4\end{gathered}$$

Since $0.5x$ means $0.5 \cdot x$, our next step is to divide each side of the equation by 0.5 to get the x by itself on one side of the equation.

$$\begin{aligned}0.5x &= 4\\\frac{0.5x}{0.5} &= \frac{4}{0.5}\\1x &= 8\\x &= 8\end{aligned}$$

The value of x is 8.

Exactly, the trickiest part is to remember the rules for adding, subtracting, multiplying and dividing decimals. Once you remember those rules, you can apply the rules to working with the equations themselves.

Example

Solve for x: $0.1(z-4.2) = 0.48$

First you can see that we have parentheses in this equation. Apply the distributive property to the left side of the equation. Multiply each of the two numbers inside the parentheses by 0.1 and then subtract those products.

$$\begin{aligned} 0.1(z-4.2) &= 0.48 \\ (0.1\times z)-(0.1\times 4.2) &= 0.48 \\ 0.1z-0.42 &= 0.48 \end{aligned}$$

Now, solve as you would solve any two-step equation. To get $0.1z$ by itself on one side of the equation, add 0.42 to both sides.

$$\begin{aligned} 0.1z-0.42 &= 0.48 \\ 0.1z-0.42+0.42 &= 0.48+0.42 \\ 0.1z+(-0.42+0.42) &= 0.9 \\ 0.1z+0 &= 0.9 \\ 0.1z &= 0.9 \end{aligned}$$

To get z by itself on one side of the equation, divide both sides by 0.1.

$$\begin{aligned} 0.1z &= 0.9 \\ \frac{0.1z}{0.1} &= \frac{0.9}{0.1} \\ 1z &= 9 \\ z &= 9 \end{aligned}$$

The value of z is 9.

Now that you know how to solve equations with decimals, let's look at solving equations involving fractions.

II. **<u>Solve Multi – Step Equations Involving Fractions</u>**

You will use the same strategy to solve a multi-step equation involving decimals that you would use to solve a multi-step equation involving integers or decimals. You will first combine like terms or use the distributive property to simplify the equation. Then, you will use inverse operations to isolate the variable on one side of the equation.

Think back. You will need to remember how to work with fractions and different operations in order to solve these equations.

Example

Solve for n: $n - \frac{n}{2} - \frac{1}{12} = \frac{5}{6}$

First, subtract the like terms n and $\frac{n}{2}$ on the left side of the equation. It may help to remember that $\frac{n}{2} = \frac{1}{2}n$ and that $n = 1n = \frac{2}{2}n$.

$$\begin{aligned} n - \frac{n}{2} - \frac{1}{12} &= \frac{5}{6} \\ \left(\frac{2}{2}n - \frac{1}{2}n\right) - \frac{1}{12} &= \frac{5}{6} \\ \frac{n}{2} - \frac{1}{12} &= \frac{5}{6} \end{aligned}$$

The next step is to isolate the term with the variable, $\frac{n}{2}$, on one side of the equation. Since $\frac{1}{12}$ is subtracted from $\frac{n}{2}$, you should add $\frac{1}{12}$ to both sides of the equation. In doing so, you will need to add $\frac{1}{12}$ and $\frac{5}{6}$, two fractions with unlike denominators. Before you add those fractions, you will need to give them a common denominator. That means you will need to find a common multiple of those two denominators and rewrite each fraction as an equivalent fraction with that denominator. Since the least common multiple of 12 and 6 is 12, you will need to rewrite $\frac{5}{6}$ as an equivalent fraction with a denominator of 12. You do not need to rewrite $\frac{1}{12}$ since it already has a denominator of 12.

$$\begin{aligned}
\frac{n}{2}-\frac{1}{12}&=\frac{5}{6}\\
\frac{n}{2}-\frac{1}{12}+\frac{1}{12}&=\frac{5}{6}+\frac{1}{12}\\
\frac{n}{2}+\left(-\frac{1}{12}+\frac{1}{12}\right)&=\frac{5}{6}+\frac{1}{12} \qquad\qquad \frac{5}{6}=\frac{5\times 2}{6\times 2}=\frac{10}{12}\\
\frac{n}{2}+0&=\frac{10}{12}+\frac{1}{12}\\
\frac{n}{2}&=\frac{11}{12}
\end{aligned}$$

Since $\frac{n}{2}$ means $n \div 2$, we should multiply each side of the equation by 2, or $\frac{2}{1}$, to get n by itself on one side of the equation.

$$\begin{aligned}
\frac{n}{2}&=\frac{11}{12}\\
\frac{n}{\cancel{2}}\times\frac{\cancel{2}}{1}&=\frac{11}{12}\times\frac{2}{1}\\
\frac{n}{1}&=\frac{22}{12}\\
n&=\frac{11}{6}=1\frac{5}{6}
\end{aligned}$$

The value of n is $1\frac{5}{6}$.

Some equations with fractions will also have a set of parentheses in them. To work with these problems, you will need to use the distributive property to simplify the equation.

Example

Solve for r: $\frac{2}{3}(r+\frac{3}{5})=2$

Apply the distributive property to the left side of the equation. Multiply each of the two numbers inside the parentheses by $\frac{2}{3}$ and then add those products.

$$\begin{aligned}
\frac{2}{3}\left(r+\frac{3}{5}\right)&=2\\
\left(\frac{2}{3}\times r\right)+\left(\frac{2}{\cancel{3}}\times\frac{\cancel{3}}{5}\right)&=2\\
\frac{2}{3}r+\frac{2}{5}&=2
\end{aligned}$$

Now, solve as you would solve any two-step equation. To get the term with the variable, $\frac{2}{3}r$, by itself on one side of the equation, subtract $\frac{2}{5}$ from both sides. To do this, it will help to rename 2 as $\frac{10}{5}$.

$$\begin{aligned}\frac{2}{3}r+\frac{2}{5}&=2\\ \frac{2}{3}r+\left(\frac{2}{5}-\frac{2}{5}\right)&=2-\frac{2}{5}\\ \frac{2}{3}r+0&=\frac{10}{5}-\frac{2}{5}\\ \frac{2}{3}r&=\frac{8}{5}\end{aligned}$$

Since $\frac{2}{3}r$ means $\frac{2}{3}\times r$, use the inverse of multiplication—division—and divide both sides of the equation by $\frac{2}{3}$. This will involve dividing $\frac{2}{3}r\div\frac{2}{3}$ on the left side of the equation. Remember, to divide two fractions, take the reciprocal of the divisor (the second fraction) and multiply that reciprocal by the dividend (the first fraction). So, $\frac{2}{3}r\div\frac{2}{3}r\times\frac{3}{2}$. Since you will be multiplying the left side of the equation by the reciprocal of $\frac{2}{3}$, which is $\frac{3}{2}$, you will need to multiply the right side of the equation by $\frac{3}{2}$ also.

$$\begin{aligned}\frac{2}{3}r&=\frac{8}{5}\\ \frac{2}{3}r\div\frac{2}{3}&=\frac{8}{5}\div\frac{2}{3}\\ \frac{2}{3}r\times\frac{3}{2}&=\frac{8}{5}\times\frac{3}{2}\\ \frac{\cancel{2}}{\cancel{3}}r\times\frac{\cancel{3}}{\cancel{2}}&=\frac{24}{10}\\ 1r&=\frac{12}{5}\\ r&=2\frac{2}{5}\end{aligned}$$

The value of r is $2\frac{2}{5}$.

Sometimes you will have equations with more than one type of rational number in the same equation. Let's look at how we can work on solving these equations.

III. <u>**Solving Multi-Step Equations Involving Multiple Forms of Rational Numbers**</u>

You know that rational numbers include integers, fractions, and terminating decimals. Some equations may require you to work with a combination of these kinds of numbers.

Example

Solve for b: $-6\left(1-\frac{b}{12}\right)=\frac{2}{3}$

This problem involves two different kinds of rational numbers: integers (-6 and 1) and fractions $\left(\frac{b}{12}\text{ and }\frac{2}{3}\right)$. You will need to know how to compute with fractions as well as how to compute with integers in order to solve this.

Apply the distributive property to the left side of the equation. Multiply each of the two numbers inside the parentheses by -6 and then subtract those products.

$$\begin{aligned}
-6\left(1-\frac{b}{12}\right) &= \frac{2}{3}\\
(-6\times 1)-\left(-6\times\frac{b}{12}\right) &= \frac{2}{3}\\
-6-\left(\frac{-6}{1}\times\frac{1}{12}b\right) &= \frac{2}{3}\\
-6-\left(\frac{-6}{12}b\right) &= \frac{2}{3}\\
-6-\left(-\frac{6}{12}b\right) &= \frac{2}{3}\\
-6+\left(\frac{6}{12}b\right) &= \frac{2}{3}
\end{aligned}$$

You may recognize immediately that the variable term, $\frac{6}{12}b$, could be simplified as $\frac{1}{2}b$ or $\frac{b}{2}$. You can wait until the problem is finished before simplifying, but if you recognize this fact, it makes sense to simplify that now. It will only make the computation easier, so simplify the variable term as $\frac{1}{2}b$.

$$\begin{aligned}
-6+\frac{6}{12}b &= \frac{2}{3}\\
-6+\frac{1}{2}b &= \frac{2}{3}
\end{aligned}$$

Now, we can solve as we would solve any two-step equation. To get $\frac{1}{2}b$ by itself on one side of the equation, we can subtract -6 from both sides.

$$\begin{aligned}
-6+\frac{1}{2}b &= \frac{2}{3}\\
-6-(-6)+\frac{1}{2}b &= \frac{2}{3}-(-6)\\
-6+6+\frac{1}{2}b &= \frac{2}{3}+6\\
0+\frac{1}{2}b &= 6\frac{2}{3}\\
\frac{1}{2}b &= 6\frac{2}{3}
\end{aligned}$$

To get b by itself, you will need to divide each side of the equation by $\frac{1}{2}$. Remember, that is the same as multiplying each side by $\frac{2}{1}$. Also, keep in mind that you will need to rewrite the mixed number $6\frac{2}{3}$ as an improper fraction $\left(\frac{20}{3}\right)$ before multiplying by $\frac{2}{1}$.

$$
\begin{aligned}
\frac{1}{2}b &= 6\frac{2}{3} \\
\frac{1}{2}b \times \frac{2}{1} &= 6\frac{2}{3} \times \frac{2}{1} \\
\frac{\cancel{1}}{\cancel{2}}b \times \frac{\cancel{2}}{\cancel{1}} &= \frac{20}{3} \times \frac{2}{1} \\
1b &= \frac{40}{3} \\
b &= 13\frac{1}{3}
\end{aligned}
$$

The value of b is $13\frac{1}{3}$.

Now, let's solve an algebraic equation that includes both decimals and fractions.

Example

Solve for k: $0.4k + 0.2k + \frac{3}{10} = \frac{9}{10}$.

First, add the like terms $0.4k$ and $0.2k$ on the left side of the equation.

$$
\begin{aligned}
0.4k + 0.2k + \frac{3}{10} &= \frac{9}{10} \\
0.6k + \frac{3}{10} &= \frac{9}{10}
\end{aligned}
$$

The next step is to isolate the term with the variable, $0.6k$, on one side of the equation. We can do this by subtracting $\frac{3}{10}$ from both sides of the equation.

$$
\begin{aligned}
0.6k + \frac{3}{10} &= \frac{9}{10} \\
0.6k + \frac{3}{10} - \frac{3}{10} &= \frac{9}{10} - \frac{3}{10} \\
0.6k + 0 &= \frac{6}{10} \\
0.6k &= \frac{6}{10}
\end{aligned}
$$

Since $0.6k$ means $0.6 \times k$, we should divide each side of the equation by 0.6 to get the k by itself on one side of the equation. This will involve dividing a fraction, $\frac{6}{10}$, by a decimal, 0.6. To do this, you will need to convert both numbers to the same form. One way to do this would be to convert the fraction $\frac{6}{10}$ to a decimal. $\frac{6}{10}$ is read as "six tenths," so the decimal form of $\frac{6}{10}$ is 0.6.

$$
\begin{aligned}
0.6k &= \frac{6}{10} \\
0.6k &= 0.6 \\
\frac{0.6k}{0.6} &= \frac{0.6}{0.6} \\
1k &= 1 \\
k &= 1
\end{aligned}
$$

The value of k is 1.

Now let's look at how these equations can be helpful when working with problems in the real world.

IV. **Model and Solve Real – World Problems Using Multi – Step Equations Involving Rational Numbers**

Sometimes, you can write a multi-step equation to represent a problem that involves rational numbers. If so, you can solve the problem by solving that equation.

Example

For a long-distance call, Guillermo's phone company charges $0.10 for the first minute and $0.05 for each minute after that. Guillermo was charged $1.00 for a long distance call he made last Friday.

a. Write an algebraic equation that could be used to represent m, the length in minutes of Guillermo's $1.00 long-distance call.

b. Determine how many minutes his $1.00 long-distance call lasted.

Consider part *a* first.

You know that the phone company charges $0.10 for the first minute and $0.05 for each minute after that. How could you represent that? If the company charged $0.05 for each minute the call lasted, you could represent that as $0.05 \times m$. However, the company charges $0.10 for the first minute and $0.05 for each minute *after* that first minute.

So, a 1-minute call will cost: $\$0.10 + (\$0.05 \times 0) = \$0.10 + \$0.00 = \$0.10$.

A 2-minute call will cost: $\$0.10 + (\$0.05 \times 1) = \$0.10 + \$0.05 = \$0.15$.

A 3-minute call will cost: $\$0.10 + (\$0.05 \times 2) = \$0.10 + \$0.10 = \$0.20$.

Notice that the number you multiply by $0.05 is always 1 less than the length of the call, in minutes. If m represents the length of a call in minutes, then this could be represented as: $\$0.10 + \$0.05 \times (m-1)$.

Write an equation that could be used to represent the cost of Guillermo's $1.00 call.

$$\begin{array}{ccccc} (\text{cost of first minute}) & + & (\text{cost of each minute after first minute}) & = & (\text{total cost}) \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 0.10 & + & 0.05(m-1) & = & 1.00 \end{array}$$

So, the equation $0.10 + 0.05(m-1) = 1.00$ represents the number of minutes that Guillermo's $1.00 phone call lasted.

Next, consider part *b*.

To find the length of the $1.00 call in minutes, solve the equation for m. First, apply the distributive property to the right side of the equation.

$$\begin{aligned} 0.10+0.05(m-1) &= 1.00 \\ 0.10+(0.05\times m)-(0.05\times 1) &= 1.00 \\ 0.10+0.05m-0.05 &= 1.00 \end{aligned}$$

Use the commutative property to rearrange the terms being added so it is easier to see how to combine the like terms. Then combine the like terms.

$$\begin{aligned} (0.10+0.05m)-0.05 &= 1.00 \\ (0.05m+0.10)-0.05 &= 1.00 \\ 0.05m+(0.10-0.05) &= 1.00 \\ 0.05m+0.05 &= 1.00 \end{aligned}$$

Now, solve as you would solve any two-step equation. First, subtract 0.05 from both sides of the equation.

$$\begin{aligned} 0.05m+0.05 &= 1.00 \\ 0.05m+0.05-0.05 &= 1.00-0.05 \\ 0.05m+0 &= 0.95 \\ 0.05m &= 0.95 \end{aligned}$$

Next, divide both sides of the equation by 0.05.

$$\begin{aligned} 0.05m &= 0.95 \\ \frac{0.05m}{0.05} &= \frac{0.95}{0.05} \\ 1m &= 19 \\ m &= 19 \end{aligned}$$

The value of m is 19, so the $1.00 call lasted for 19 minutes.

Now let's take a look at a problem that has fractions in it.

Example

On Sunday, Leah walked 4 miles. On Monday, Leah walked one-third as many miles as she walked on Tuesday. She walked a total of 12 miles on those 3 days.

a. Let t represent the number of miles Leah walked today. Write an algebraic equation to represent the total number of miles she walked on all 3 days.

b. Find the number of miles Leah walked on Tuesday.

c. Find the number of miles Leah walked on Monday.

Consider part *a* first.

You know that t represents the number of miles Leah walked on Tuesday. Use that variable to write an expression for the number of miles Leah walked on Monday.

$$\begin{array}{c} On\ Monday,\ Leah\ walked\ \underline{one-third\ as\ many\ miles\ as\ldots on\ Tuesday}. \\ \downarrow \\ \frac{t}{3}\ or\ \frac{1}{3}t \end{array}$$

So, you know that Leah walked 4 miles on Sunday, t miles on Monday, and $\frac{1}{3}t$ miles on Tuesday. You also know that she walked a *total* of 12 miles on all three days. Use this information to write an addition equation for this problem.

$$\begin{array}{ccccccc} (\text{miles walked Sun.}) & + & (\text{miles walked Mon.}) & + & (\text{miles walked Tues.}) & = & (\text{total miles walked}) \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 4 & + & \frac{1}{3}t & + & t & = & 12 \end{array}$$

So, this problem can be represented by the equation, $4+\frac{1}{3}t+t=12$.

Next, consider part *b*.

The variable t represents the number of miles Leah walked today. So, solve the equation for t. Start by adding the like terms on the left side of the equation.

$$\begin{aligned}4+\frac{1}{3}t+t&=12\\4+\frac{1}{3}t+\frac{3}{3}t&=12\\4+\frac{4}{3}t&=12\end{aligned}$$

Solve the equation for t as you would solve any two-step equation. Subtract 4 from both sides of the equation.

$$\begin{aligned}4+\frac{4}{3}t&=12\\4-4+\frac{4}{3}t&=12-4\\0+\frac{4}{3}t&=8\\\frac{4}{3}t&=8\end{aligned}$$

Finally, you must divide both sides of the equation by $\frac{4}{3}$. Remember, that is the same as multiplying both sides of the equation by $\frac{3}{4}$.

$$\begin{aligned}\frac{4}{3}t&=8\\\frac{4}{3}t\times\frac{3}{4}&=8\times\frac{3}{4}\\\frac{\cancel{4}}{\cancel{3}}t\times\frac{\cancel{3}}{\cancel{4}}&=\frac{8}{1}\times\frac{3}{4}\\1t&=\frac{24}{4}\\t&=6\end{aligned}$$

The value of t is 6, so Leah walked 6 miles on Tuesday.

Consider part c next.

In part a, you determined that Leah walked $\frac{1}{3}t$ miles on Monday. Since $t=6$, substitute 6 for t in the expression to find how many miles she walked yesterday.

$$\frac{1}{3}t=\frac{1}{3}\times 6=\frac{1}{3}\times\frac{6}{1}=\frac{6}{3}=2$$

Leah walked 2 miles on Monday.

That may seem like a lot of steps, but if you just work through each part of the problem, you will be able to solve it. Now let's go back and solve the problem from the introduction.

Real-Life Example Completed

The New Piece of Music

Here is the original problem once again. Remember that you are trying to figure out how long Jose practiced on Sunday and Monday. To do this, write an equation. Then solve it for the two times. You will have three parts to your answer.

Jose has been playing the French horn for many years and until now, everything has been easy. Now Mrs. Kline, the band director has assigned him a new piece of music to work on and it is very tricky. Jose has been practicing the new piece.

His best rehearsal was on Saturday, when he practiced for 90 minutes. On Sunday, he had a birthday party to go to, so he did not practice as long. On Monday, he had a math test to study for and so he practiced for half as long as he did on Monday.

When Jose went to band practice on Tuesday afternoon, he struggled through the piece.

"How long did you practice?" Mrs. Kline asked him.

"Well, from Saturday to Tuesday I practiced a total of 3 hours," Jose said.

Now write an equation and solve for the missing times. Remember there are three parts to your answer.

Solution to the Real – Life Example

First, write an equation to show what you know and what you don't know.

Saturday = 90 minutes

Monday $= t -$ **missing time**

Tuesday $= \frac{1}{2}t -$ **half the time of Monday**

Total time = 3 hours

$90 + t + \frac{1}{2}t = 3 \ hours$

First, convert hours to minutes.

$90 + t + \frac{1}{2}t = 180 \ minutes$

Now we can solve the equation.

Monday's time = 60 minutes

Tuesday's time = 30 minutes

Vocabulary

Here are the vocabulary words that are found in this lesson.

Integer
the set of whole numbers and their opposites.

Rational Numbers
a set of numbers that includes integers, decimals, fractions, terminating and repeating decimals. These numbers can be written in fraction form.

Fraction
a part of a whole written using a numerator and a denominator.

Decimal
a part of a whole written using place value and a decimal point.

Repeating Decimal
a decimal where the digits repeat in a pattern and eventually end.

Terminating Decimal
a decimal where the digits eventually end, but where numbers do not repeat in a pattern.

Time to Practice

Directions: Solve each equation to find the value of the variable.

1. $7n - 3.2n + 6.5 = 14.1$
2. $0.2(3 + p) = 5.6$
3. $s + \frac{s}{5} + \frac{2}{3} = \frac{5}{6}$
4. $j - \frac{j}{7} - \frac{1}{2} = 11\frac{1}{2}$
5. $\frac{3}{4}\left(g - \frac{1}{2}\right) = \frac{1}{8}$
6. $-2\left(1 - \frac{a}{4}\right) = \frac{1}{8}$
7. $0.09y - 0.08y + \frac{1}{10} = \frac{3}{10}$
8. $.06x + .05x = .99$
9. $\frac{1}{2}y + \frac{1}{3}y = 24$
10. $\frac{1}{4}x + \frac{1}{3} = \frac{2}{3}$
11. $\frac{1}{2}x = 18$
12. $.9x = 56$
13. $.6x + 1 = 19$
14. $\frac{1}{4}x + 2 = 19$
15. $9.05x = 27.15$

Directions: Solve each problem.

Maggie went to the mall with d dollars in her wallet. She spent half of the money in her wallet at the music store. Then she spent another \$6 at the book store. After going to both of those stores, she had \$9 left in her wallet.

16. Remember, d represents the number of dollars in Maggie's wallet when she went to the mall, before she purchased anything. Write an algebraic equation to represent this problem.
17. Find d, the number of dollars Maggie had in her wallet when she went to the mall.
18. Find the number of dollars Maggie spent at the music store. The cost of a taxi ride depends on the number of miles driven. The rider is charged \$4.50 for the first mile and \$2.50 for each mile after that. Mr. O'Connor paid \$12.00 for a taxi ride last week.
19. Write an algebraic equation that could be used to represent m, the number of miles Mr. O'Connor traveled during his \$12.00 taxi ride.

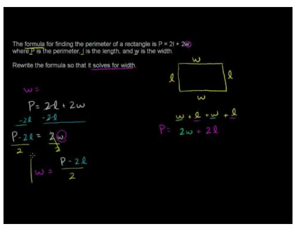
20. Determine how many miles Mr. O'Connor traveled during his \$12.00 taxi ride. Each week, Darius saves some of his earnings and puts it in a bank. During week 1, Darius saved x dollars. During week 2, Darius saved two-thirds as many dollars as he had during week 1. During week 3, Darius saved \$10.
21. Darius saved a total of \$60 during weeks 1, 2, and 3. Write an algebraic equation to represent x, the number of dollars Darius saved during week 1.
22. Find the number of dollars Darius saved during week 1.
23. Find the number of dollars Darius saved during week 2.

3.5 Solving Algebraic Equations for a Variable

Here you'll learn how to isolate the variable in an equation or formula.

You are planning a trip to Spain in the summer. In the U.S., we use Fahrenheit to measure the temperature, but in Europe, they use Celsius. To get yourself used to Celsius, you find the equation $F = \frac{9}{5}C + 32$. You then substitute the temperature in Celsius for C and solve the equation for F. When you start looking at the weather in Spain in the summer, the website says it is usually $35^\circ C$. What should you pack?

Watch This



MEDIA

Click image to the left for more content.

Khan Academy: Example of Solving for a Variable

Guidance

Solving an algebraic equation for a variable can be tricky, but it can also be very useful. This technique can be used to go back and forth between to different units of measurement. To solve for, or **isolate**, a variable within an equation, you must undo the operations that are in the equation.

Example A

Solve for x in the equation $3x - 4y = 12$.

Solution: To solve for x, you need to move the $4y$ to the other side of the equation. In order to do this, we must do the opposite operation. Since the $4y$ is being subtracted, we must add it to the other side of the equation.

$$\begin{aligned} 3x - 4y &= 12 \\ +4y &= +4y \\ \hline 3x &= 4y + 12 \end{aligned}$$

Now we need to get x alone. $3x$ is the same as "3 multiplied by x." Therefore, to undo the multiplication we must divide both sides of the equation by 3.

$$\begin{aligned} \frac{3x}{3} &= \frac{4y}{3} + \frac{12}{3} \\ x &= \frac{4}{3}y + 4 \end{aligned}$$

When undoing multiplication, you must divide *everything* in the equation by 3.

A few things to note:

1. To undo an operation within an equation, do the opposite.
2. Perform the opposite operations in the reverse order of the Order of Operations.
3. Combine all like terms on the same side of the equals sign before doing #1.

Example B

Given the equation $\frac{2b}{3}+6a-4=8$, find b when $a=1$ and $a=-2$.

Solution: This example combines what was learned in the last section with what we did in the previous example. First, isolate b. Move both the $6a$ and the 4 over to the other side by doing the opposite operation.

$$\begin{aligned}
\frac{2b}{3}+6a-4&=8\\
\underline{\quad -6a+4}&\underline{=+4-6a}\\
\frac{2b}{3}&=12-6a
\end{aligned}$$

Now, we have a fraction multiplied by b. To undo this we must multiply by the reciprocal of $\frac{2}{3}$, which is $\frac{3}{2}$. This means, we must multiply everything in the equation by the reciprocal.

$$\begin{aligned}
\frac{2b}{3}&=12-6a\\
\frac{3}{2}\left(\frac{2b}{3}\right.&\left.=12-6a\right)\\
\frac{3}{2}\cdot\frac{2b}{3}&=\frac{3}{2}\cdot 12-\frac{3}{2}\cdot 6a\\
b&=18-9a
\end{aligned}$$

Even though we know that a 9 can be pulled out of the 18 and the 9 in the above equation, it is not a necessary step in solving for b. Now, we can do what the question asks, find b when $a=1$ and -2. Plug in these values for a.

$$a=1:b=18-9(1)=18-9=9 \qquad\qquad a=-2:18-9(-2)=18+18=36$$

Example C

The area of a triangle is $A=\frac{1}{2}bh$, where b is the base of the triangle and h is the height. You know that the area of a triangle is 60 in^2 and the base is 12 in. Find the height.

Solution: First solve the equation for h.

$$
\begin{aligned}
A &= \frac{1}{2}bh \\
2 \cdot A &= 2 \cdot \frac{1}{2}bh && \text{Multiply both sides by 2.} \\
2A &= bh && \text{Divide both side by } b. \\
\frac{2A}{b} &= h
\end{aligned}
$$

Now, plug in what we know to find h. $\frac{2(60)}{12} = \frac{120}{12} = 10$. The height is 10 inches.

This process is helpful when solving for any variable in any equation or formula. Some formulas you may need to know for this text are listed below.

TABLE 3.1:

Distance	$d = rt$	$d =$ distance, $r =$ rate, $t =$ time
Temperature	$F = \frac{9}{5}C + 32$	$F =$ degrees in Fahrenheit, $C =$ degrees in Celsius
Area of a Triangle	$A = \frac{1}{2}bh$	$A =$ area, $b =$ base, $h =$ height
Area of a Rectangle	$A = bh$	$A =$ area, $b =$ base, $h =$ height
Area of a Circle	$A = \pi r^2$	$A =$ area, $r =$ radius
Area of a Trapezoid	$A = \frac{1}{2}h(b_1 + b_2)$	$A =$ area, $h =$ height, $b_1 =$ one base, $b_2 =$ other base
Perimeter of a Rectangle	$P = 2l + 2w$	$P =$ perimeter, $l =$ length, $w =$ width
Circumference of a Circle	$C = 2\pi r$	$C =$ circumference, $r =$ radius

Intro Problem Revisit First, let's substitute in $C = 35^\circ$ to find the degrees in Fahrenheit.

$$
\begin{aligned}
F &= \frac{9}{5} \cdot 35 + 32 \\
F &= 63 + 32 \\
F &= 95^\circ
\end{aligned}
$$

As you can see, a good estimator for the conversion is to double the degrees, when in Celsius, and then add 32. Needless to say, it is going to be quite hot in Spain and you should pack summer clothes.

Guided Practice

1. Solve $\frac{3}{4}x - 2y = 15$ for x.

2. If the temperature is $41^\circ F$, what is it in Celsius?

Answers

1. First add $2y$ to both sides, then multiply both sides by the reciprocal of $\frac{3}{4}$.

$$\begin{array}{r} \frac{3}{4}x - 2y = 15 \\ +2y \ +2y \\ \hline \frac{3}{4}x = 2y + 15 \\ x = \frac{4}{3} \cdot 2y + \frac{4}{3} \cdot 15 \\ x = \frac{8}{3}y + 20 \end{array}$$

2. You can solve this problem two different ways: First, plug in 41 for F and then solve for C or solve for C first and then plug in 41 for F. We will do the second option.

$$\begin{aligned} F &= \frac{9}{5}C + 32 \\ F - 32 &= \frac{9}{5}C \\ \frac{5(F-32)}{9} &= C \end{aligned}$$

Now, plug in 41 for F. $\frac{5(41-32)}{9} = \frac{5 \cdot 9}{9} = 9^\circ C$.

Practice

Solve the following equations or formulas for the indicated variable.

1. $6x - 3y = 9$; solve for y.
2. $4c + 9d = 16$; solve for c.
3. $5f - 6g = 14$; solve for f.
4. $\frac{1}{3}x + 5y = 1$; solve for x.
5. $\frac{4}{5}m + \frac{2}{3}n = 24$; solve for m.
6. $\frac{4}{5}m + \frac{2}{3}n = 24$; solve for n.
7. $P = 2l + 2w$; solve for w.
8. $F = \frac{9}{5}C + 32$; solve for C.

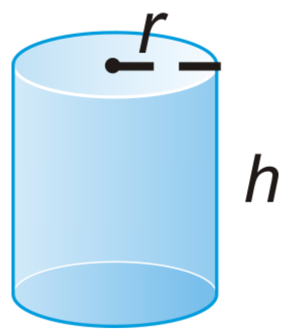
Find the value of y given the value of x.

9. $4x - 8y = 2; x = -1$
10. $2y - 5x = 12; x = 16$
11. $3x + \frac{1}{2}y = -5; x = 7$
12. $\frac{1}{4}x + \frac{2}{3}y - 18 = 0; x = -24$

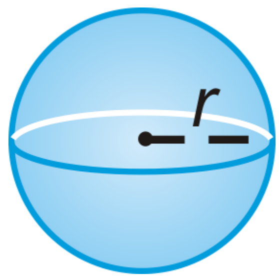
For questions 13-15, use the formulas in the chart above to answer the following questions.

13. If the temperature is $86^\circ F$, what is it in Celsius?
14. If the area of a circle is $36\pi\ cm^2$, what is the radius?
15. The area of a trapezoid is $72\ ft^2$, the height is 8 ft and b_1 is 6, what is the length of the other base?

For questions 16-17, use the equation for the surface area of a cylinder, $SA = 2\pi r^2 + 2\pi rh$, where r is the radius and h is the height.



16. Solve the equation for h.
17. Find h if the surface area is $120\pi\ cm^2$ and the radius is 6 cm.
18. **Challenge** The formula for the volume of a sphere is $V = \frac{4}{3}\pi r^3$. Solve the equation for r, the radius.

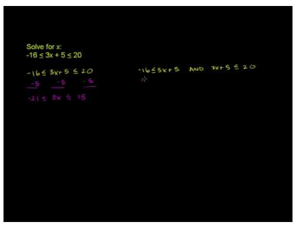


3.6 Compound Inequalities

Here you will solve two inequalities that have been joined together by the words “and” and “or.”

Mr. Garcia, the Spanish teacher, announces that the students’ final grades will consist of 40% projects (0-100 score). The remaining 60% (0-100) will come from the final exam. Going into the final exam, Madison has an 84 in the project score. Within what grade range must she fall on the final exam to raise her score to an A (90-100 overall score)?

Watch This



MEDIA

Click image to the left for more content.

Khan Academy: Compound Inequalities 3

Guidance

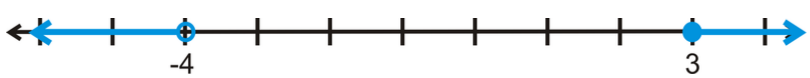
Compound inequalities are inequalities that have been joined by the words “and” or “or.” For example:

$-2 < x \leq 5$ Read, “x is greater than -2 *and* less than or equal to 5.”

$x \geq 3$ or $x < -4$ Read, “x is greater than or equal to 3 *or* less than -4.”

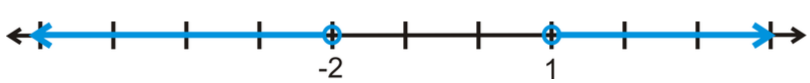
Notice that both of these inequalities have two inequality signs. So, it is like solving or graphing two inequalities at the same time. When graphing, look at the inequality to help you. The first compound inequality above, $-2 < x \leq 5$, has the x in between -2 and 5, so the shading will also be between the two numbers.

And, with the “or” statement, the shading will go in opposite directions.



Example A

Write the inequality statement given by the graph below.



Solution: Because the shading goes in opposite directions, we know this is an “or” statement. Therefore, the statement is $x < -2$ or $x > 1$.

Example B

Solve and graph $-3 < 2x+5 \leq 11$.

Solution: This is like solving two inequalities at the same time. You can split the statement apart to have two inequalities, $-3 < 2x+5$ and $2x+5 \leq 11$ and solve. You can also leave the compound inequality whole to solve.

$$\begin{aligned}
&-3 < 2x+\cancel{5} \leq 11 \\
&\underline{-5 \qquad -\cancel{5} \quad -5} \\
&\frac{-8}{2} < \frac{\cancel{2}x}{\cancel{2}} \leq \frac{6}{2} \\
&-4 < x \leq 3
\end{aligned}$$

Test a solution, $x = 0:$

$$\begin{aligned}
-3 < 2(0)+5 \leq 11 \\
-3 < 5 \leq 11
\end{aligned}$$

Here is the graph:

Example C

Solve and graph $-32 > -5x+3$ or $x-4 \leq 2$.

Solution: When solving an "or" inequality, solve the two inequalities separately, but show the solution on the same number line.

$$\begin{aligned}
&-32 > -5x+\cancel{3} \;\; or \;\; x-\cancel{4} \leq 2 \\
&\underline{-3 \qquad\quad -\cancel{3}} \qquad \underline{+\cancel{4} \; +4} \\
&\frac{-35}{-5} > \frac{-\cancel{5}x}{-\cancel{5}} \qquad\qquad x \leq 6 \\
&\quad 7 < x
\end{aligned}$$

Notice that in the first inequality, we had to flip the inequality sign because we divided by -5. Also, it is a little more complicated to test a solution for these types of inequalities. You still test one point, but it will only work for one of the inequalities. Let's test $x = 10$. First inequality: $-32 > -5(10)+3 \rightarrow -32 > -47$. Second inequality: $10-4 \leq 2 \rightarrow 5 \nleq 2$. Because $x = 10$ works for the first inequality, is a solution. Here is the graph.

Intro Problem Revisit Writing the grading as an expression, we get $0.4(84)+0.6x$ where x is the final exam score. Madison wants to get an A, so we will have a compound inequality that ranges between 90 and 100.

$$\begin{aligned}90 \leq 33.6+0.6x &\leq 100\\ 56.4 \leq 0.6x &\leq 66.4\\ 94 \leq x &\leq 110.67\end{aligned}$$

Unless Mr. Garcia offers extra credit, Madison can't score higher than 100. So, she has to score at least 94 or more, up to 100, to get an A. This represents a constraint on the inequality. Although 105 is part of the solution set, the constraints make it a non-viable solution.

Interval Notation

Compound inequalities can also be graphed using interval notation. For more information click here.

Guided Practice

1. Graph $-7 \leq x \leq -1$ on a number line.

Solve the following compound inequalities and graph.

2. $5 \leq -\frac{2}{3}x+1 \leq 15$

3. $\frac{x}{4}-7>5$ or $\frac{8}{5}x+2 \leq 18$

Answers

1. This is an "and" inequality, so the shading will be between the two numbers.

2. Solve this just like Example B.

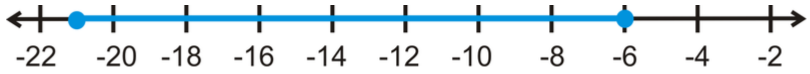
$$\begin{array}{c} 5 \leq -\frac{2}{3}x+\cancel{1} \leq 15\\ \underline{-1 \qquad\qquad -\cancel{1} \quad -1}\\ 4 \leq -\frac{2}{3}x \leq 14\\ -\frac{3}{2}\left(4 \leq -\frac{2}{3}x \leq 14\right)\\ -6 \geq x \geq -21\end{array}$$

Test a solution, $x=-10$:

$$\begin{aligned}5 \leq -\frac{2}{3}(-12)+1 &\leq 15\\ 5 \leq 9 &\leq 15\end{aligned}$$

This solution can also be written $-21 \leq x \leq -6$.

The graph is:



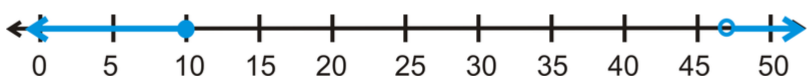
3. This is an “or” compound inequality. Solve the two inequalities separately.

$$\begin{array}{rcl} \frac{x}{4}-\cancel{7}>5 & or & \frac{8}{5}x+\cancel{2}\leq 18 \\ \underline{+\cancel{7}\ +7} & & \underline{-\cancel{2}\ \ -2} \\ \cancel{4}\cdot\frac{x}{\cancel{4}}>12\cdot 4 & or & \frac{\cancel{5}}{\cancel{8}}\cdot\frac{\cancel{8}}{\cancel{5}}x\leq 16\cdot\frac{5}{8} \\ x>48 & or & x\leq 10 \end{array}$$

Test a solution, $x=0$:

$$\begin{array}{rcl} \frac{0}{4}-7>5 & or & \frac{8}{5}(0)+2\leq 18 \\ -7 \ngtr 5 & or & 2\leq 18 \end{array}$$

Notice that $x=0$ is a solution for the second inequality, which makes it a solution for the entire compound inequality. Here is the graph:



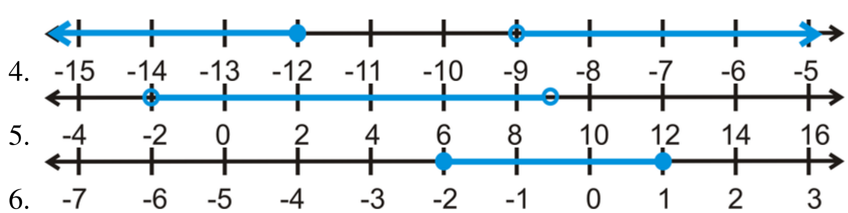
On problems 2 and 3 we changed the scale of the number line to accommodate the solution.

Practice

Graph the following compound inequalities. Use an appropriate scale.

1. $-1<x<8$
2. $x>5$ or $x\leq 3$
3. $-4\leq x\leq 0$

Write the compound inequality that best fits each graph below.



Solve each compound inequality and graph the solution using interval notation.

7. $-11<x-9\leq 2$
8. $8\leq 3-5x<28$
9. $2x-7>-13$ or $\frac{1}{3}x+5\leq 1$

10. $0 < \frac{x}{5} < 4$
11. $-4x+9 < 35$ or $3x-7 \leq -16$
12. $\frac{3}{4}x+7 \geq -29$ or $16-x > 2$
13. $3 \leq 6x-15 < 51$
14. $-20 < -\frac{3}{2}x+1 < 16$
15. **Challenge** Write a compound inequality whose solutions are all real numbers. Show why this is true.

3.7 Applications with Inequalities

Here you'll learn how to use inequalities to solve real-world problems.

Suppose that a company's budget requires it to spend at least \$20,000 but no more than \$30,000 on training for its employees. The cost of training is the combination of a flat fee of \$5,000 plus \$500 per employee. If the company has m employees, how much is it required to spend per employee? In this Concept, you'll learn how to solve real-world problems such as this one by using inequalities.

Guidance

As you saw with equations, inequalities are also useful for solving real-world problems. In this Concept, you will see some examples of how to set up and solve real-world problems.

Example A

In order to get a bonus this month, Leon must sell at least 120 newspaper subscriptions. He sold 85 subscriptions in the first three weeks of the month. How many subscriptions must Leon sell in the last week of the month?

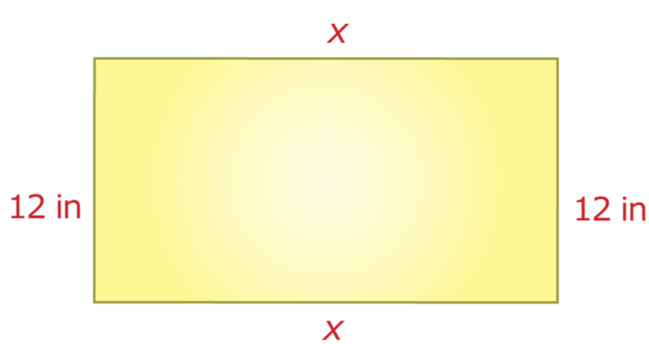
Solution: The number of subscriptions Leon needs is "at least" 120. Choose a variable, such as n, to represent the varying quantity–the number of subscriptions. The inequality that represents the situation is $n+85 \geq 120$.

Solve by isolating the variable n: $n \geq 35$.

Leon must sell 35 or more subscriptions to receive his bonus.

Example B

The width of a rectangle is 12 inches. What must the length be if the perimeter is at least 180 inches? (Note: Diagram not drawn to scale.)



Solution: The perimeter is the sum of all the sides.

$$12+12+x+x \geq 180$$

Simplify and solve for the variable x:

$$12+12+x+x \geq 180 \rightarrow 24+2x \geq 180$$
$$2x \geq 156$$
$$x \geq 78$$

The length of the rectangle must be 78 inches or greater.

Example C

The speed of a golf ball in the air is given by the formula $v=-32t+80$*, where t is the time since the ball was hit. When is the ball traveling between 20 ft/sec and 30 ft/sec inclusive?*

Solution: We want to find the times when the ball is traveling between 20 ft/sec and 30 ft/sec inclusive. Begin by writing the inequality to represent the unknown values, $20 \leq v \leq 30$.

Replace the velocity formula, $v=-32t+80$, with the minimum and maximum values.

$$20 \leq -32t+80 \leq 30$$

Separate the compound inequality and solve each separate inequality.

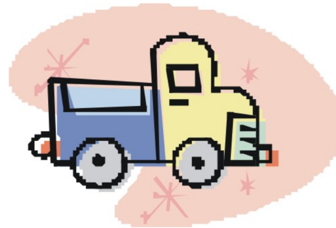
$$\begin{array}{lcr} 20 \leq -32t+80 & & -32t+80 \leq 30 \\ 32t \leq 60 & \text{and} & 50 \leq 32t \\ t \leq 1.875 & & 1.56 \leq t \end{array}$$

$1.56 \leq t \leq 1.875$. Between 1.56 and 1.875 seconds, the ball is traveling between 20 ft/sec and 30 ft/sec.

Inequalities can also be combined with dimensional analysis.

Guided Practice

William's pick-up truck gets between 18 and 22 miles per gallon of gasoline. His gas tank can hold 15 gallons of gasoline. If he drives at an average speed of 40 miles per hour, how much driving time does he get on a full tank of gas?



Solution: Use dimensional analysis to get from time per tank to miles per gallon.

$$\frac{t\ \cancel{hour}}{1\ \cancel{tank}} \times \frac{1\ \cancel{tank}}{15\ gallons} \times \frac{40\ miles}{1\ \cancel{hour}} = \frac{40t}{45}\ \frac{miles}{gallon}$$

Since the truck gets between 18 to 22 miles/gallon, you can write a compound inequality.

$$18 \le \frac{40t}{15} \le 22$$

Separate the compound inequality and solve each inequality separately.

$$\begin{aligned} 18 &\le \frac{40t}{15} \\ 270 &\le 40t \\ 6.75 &\le t \end{aligned} \qquad \text{and} \qquad \begin{aligned} \frac{40t}{15} &\le 22 \\ 40t &\le 330 \\ t &\le 8.25 \end{aligned}$$

Andrew can drive between 6.75 and 8.25 hours on a full tank of gas.

Practice

For problems 1-5, write the inequality given by the statement. Choose an appropriate letter to describe the unknown quantity.

1. You must be at least 48 inches tall to ride the "Thunderbolt" Rollercoaster.
2. You must be younger than 3 years old to get free admission at the San Diego Zoo.
3. Charlie needs more than $1,800 to purchase a car.
4. Cheryl can have no more than six pets at her house.
5. The shelter can house no more than 16 rabbits.

6. The width of a rectangle is 16 inches. Its area is greater than 180 square inches.
   (a) Write an inequality to represent this situation.
   (b) Graph the possible lengths of the rectangle.

7. Ninety percent of some number is at most 45.
   (a) Write an inequality to represent the situation.
   (b) Write the solutions as an algebraic sentence.

8. Doubling Martha's jam recipe yields at least 22 pints.

(a) Write an inequality to represent the situation.
(b) Write the solutions using interval notation.

For problems 9-15, write the inequality and use it solve the problem.

9. At the San Diego Zoo, you can either pay $22.75 for the entrance fee or $71 for the yearly pass, which entitles you to unlimited admission. At most, how many times can you enter the zoo for the $22.75 entrance fee before spending more than the cost of a yearly membership?
10. Proteek's scores for four tests were 82, 95, 86, and 88. What will he have to score on his last test to average at least 90 for the term?
11. Raul is buying ties and he wants to spend $200 or less on his purchase. The ties he likes the best cost $50. How many ties could he purchase?
12. Virena's Scout Troop is trying to raise at least $650 this spring. How many boxes of cookies must they sell at $4.50 per box in order to reach their goal?
13. Using the golf ball example, find the times in which the velocity of the ball is between 50 *ft/sec and* 60 *ft/sec*.
14. Using the pick-up truck example, suppose William's truck has a dirty air filter, causing the fuel economy to be between 16 and 18 miles per gallon. How many hours can William drive on a full tank of gas using this information?
15. To get a grade of B in her Algebra class, Stacey must have an average grade greater than or equal to 80 and less than 90. She received the grades of 92, 78, and 85 on her first three tests. Between which scores must her grade fall on her last test if she is to receive a grade of B for the class?

Mixed Review

16. Solve the inequality and write its solution in interval notation: $\frac{x+3}{2} < -4$.
17. Graph $2x-2y=6$ using its intercepts.
18. Identify the slope and $y-$intercept of $y+1=\frac{2}{5}(x-5)$.
19. George rents videos through a mail-order company. He can get 16 movies each month for $16.99. Sheri rents videos through instant watch. She pays $1.99 per movie. When will George pay less than Sheri?
20. Evaluate: $-2\frac{1}{5} \div 1\frac{3}{4}$.

3.8 Graphing Linear Inequalities in Two Variables

Here you'll learn how to graph linear inequalities in two variables of the form $y > mx + b$ or $y < mx + b$. You'll also solve real-world problems involving such inequalities.

What if you were given a linear inequality like $2x - 3y \leq 5$? How could you graph that inequality in the coordinate plane? After completing this Concept, you'll be able to graph linear inequalities in two variables like this one.

Watch This

MEDIA

Click image to the left for more content.

CK-12 Foundation: 0612S Linear Inequalities in Two Variables (H264)

Guidance

The general procedure for graphing inequalities in two variables is as follows:

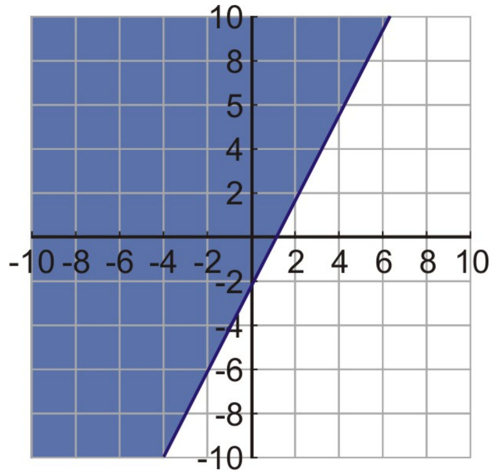
1. Re-write the inequality in slope-intercept form: $y = mx + b$. Writing the inequality in this form lets you know the direction of the inequality.
2. Graph the line of the equation $y = mx + b$ using your favorite method (plotting two points, using slope and $y-$intercept, using $y-$intercept and another point, or whatever is easiest). Draw the line as a dashed line if the equals sign is not included and a solid line if the equals sign is included.
3. Shade the half plane above the line if the inequality is "greater than." Shade the half plane under the line if the inequality is "less than."

Example A

Graph the inequality $y \geq 2x - 3$.

Solution

The inequality is already written in slope-intercept form, so it's easy to graph. First we graph the line $y = 2x - 3$; then we shade the half-plane above the line. The line is solid because the inequality includes the equals sign.



Example B

Graph the inequality $5x-2y>4$.

Solution

First we need to rewrite the inequality in slope-intercept form:

$$-2y > -5x+4$$
$$y < \frac{5}{2}x-2$$

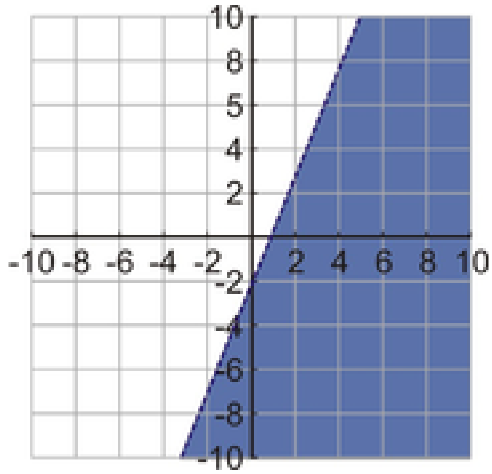
Notice that the inequality sign changed direction because we divided by a negative number.

To graph the equation, we can make a table of values:

TABLE 3.2:

x	y
-2	$\frac{5}{2}(-2)-2=-7$
0	$\frac{5}{2}(0)-2=-2$
2	$\frac{5}{2}(2)-2=3$

After graphing the line, we shade the plane **below** the line because the inequality in slope-intercept form is **less than**. The line is dashed because the inequality does not include an equals sign.



Solve Real-World Problems Using Linear Inequalities

In this section, we see how linear inequalities can be used to solve real-world applications.

Example C

A retailer sells two types of coffee beans. One type costs \$9 per pound and the other type costs \$7 per pound. Find all the possible amounts of the two different coffee beans that can be mixed together to get a quantity of coffee beans costing \$8.50 or less.

Solution

Let $x =$ weight of \$9 per pound coffee beans in pounds.

Let $y =$ weight of \$7 per pound coffee beans in pounds.

The cost of a pound of coffee blend is given by $9x+7y$.

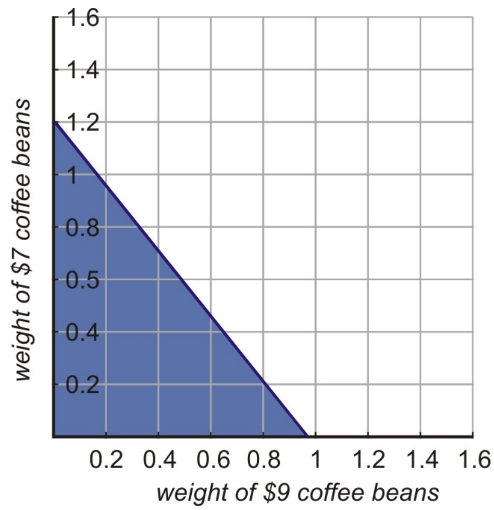
We are looking for the mixtures that cost \$8.50 or less. We write the inequality $9x+7y \le 8.50$.

Since this inequality is in standard form, it's easiest to graph it by finding the $x-$ and $y-$intercepts. When $x = 0$, we have $7y = 8.50$ or $y = \frac{8.50}{7} \approx 1.21$. When $y = 0$, we have $9x = 8.50$ or $x = \frac{8.50}{9} \approx 0.94$. We can then graph the line that includes those two points.

Now we have to figure out which side of the line to shade. In $y-$intercept form, we shade the area **below** the line when the inequality is "less than." But in standard form that's not always true. We could convert the inequality to $y-$intercept form to find out which side to shade, but there is another way that can be easier.

The other method, which works for any linear inequality in any form, is to plug a random point into the inequality and see if it makes the inequality true. Any point that's not on the line will do; the point (0, 0) is usually the most convenient.

In this case, plugging in 0 for x and y would give us $9(0)+7(0) \le 8.50$, which is true. That means we should shade the half of the plane that includes (0, 0). If plugging in (0, 0) gave us a false inequality, that would mean that the solution set is the part of the plane that does *not* contain (0, 0).



Notice also that in this graph we show only the first quadrant of the coordinate plane. That's because weight values in the real world are always nonnegative, so points outside the first quadrant don't represent real-world solutions to this problem.

Watch this video for help with the Examples above.

MEDIA

Click image to the left for more content.

CK-12 Foundation: Linear Inequalities in Two Variables

Vocabulary

- For a strict inequality, we draw a **dashed line** to show that the points in the line *are not* part of the solution. For an inequality that includes the equals sign, we draw a **solid line** to show that the points on the line *are* part of the solution.

- The solution to a linear inequality includes all the points in one half of the plane. We can tell which half by looking at the inequality sign:

$>$The solution set is the half plane above the line.

$\geq$ The solution set is the half plane above the line and also all the points on the line.

$<$The solution set is the half plane below the line.

$\leq$ The solution set is the half plane below the line and also all the points on the line.

Guided Practice

Julius has a job as an appliance salesman. He earns a commission of $60 for each washing machine he sells and $130 for each refrigerator he sells. How many washing machines and refrigerators must Julius sell in order to make $1000 or more in commissions?

Solution

Let $x =$ number of washing machines Julius sells.

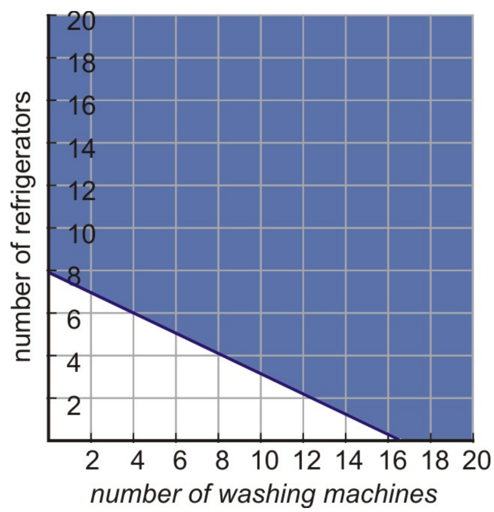
Let $y =$ number of refrigerators Julius sells.

The total commission is $60x + 130y$.

We're looking for a total commission of \$1000 or more, so we write the inequality $60x + 130y \geq 1000$.

Once again, we can do this most easily by finding the $x-$ and $y-$intercepts. When $x = 0$, we have $130y = 1000$, or $y = \frac{1000}{30} \approx 7.69$. When $y = 0$, we have $60x = 1000$, or $x = \frac{1000}{60} \approx 16.67$.

We draw a solid line connecting those points, and shade above the line because the inequality is "greater than." We can check this by plugging in the point (0, 0): selling 0 washing machines and 0 refrigerators would give Julius a commission of \$0, which is *not* greater than or equal to \$1000, so the point (0, 0) is *not* part of the solution; instead, we want to shade the side of the line that does *not* include it.



Notice also that we show only the first quadrant of the coordinate plane, because Julius's commission should be non-negative.

Practice

Graph the following inequalities on the coordinate plane.

1. $y \leq 4x + 3$
2. $y > -\frac{x}{2} - 6$
3. $3x - 4y \geq 12$
4. $x + 7y < 5$
5. $6x + 5y > 1$
6. $y + 5 \leq -4x + 10$
7. $x - \frac{1}{2}y \geq 5$
8. $6x + y < 20$
9. $30x + 5y < 100$
10. Remember what you learned in the last chapter about families of lines.
    a. What do the graphs of $y > x + 2$ and $y < x + 5$ have in common?
    b. What do you think the graph of $x + 2 < y < x + 5$ would look like?
11. How would the answer to problem 6 change if you subtracted 2 from the right-hand side of the inequality?

12. How would the answer to problem 7 change if you added 12 to the right-hand side?
13. How would the answer to problem 8 change if you flipped the inequality sign?
14. A phone company charges 50 cents per minute during the daytime and 10 cents per minute at night. How many daytime minutes and nighttime minutes could you use in one week if you wanted to pay less than $20?
15. Suppose you are graphing the inequality $y > 5x$.
    a. Why can't you plug in the point (0, 0) to tell you which side of the line to shade?
    b. What happens if you do plug it in?
    c. Try plugging in the point (0, 1) instead. Now which side of the line should you shade?
16. A theater wants to take in at least $2000 for a certain matinee. Children's tickets cost $5 each and adult tickets cost $10 each.
    a. If x represents the number of adult tickets sold and y represents the number of children's tickets, write an inequality describing the number of tickets that will allow the theater to meet their minimum take.
    b. If 100 children's tickets and 100 adult tickets have already been sold, what inequality describes how many *more* tickets of both types the theater needs to sell?
    c. If the theater has only 300 seats (so only 100 are still available), what inequality describes the *maximum* number of additional tickets of both types the theater can sell?

Texas Instruments Resources

In the CK-12 Texas Instruments Algebra I FlexBook, there are graphing calculator activities designed to supplement the objectives for some of the lessons in this chapter. See http://www.ck12.org/flexr/chapter/9616.

Chapter 4 Systems of Linear Equations and Inequalities

Chapter Outline

4.1 Graphs of Linear Systems

Here you'll learn how to use a graph to solve a system of linear equations.

Suppose that on a test there are multiple choice questions and fill-in-the-blank questions. Julia answered 8 multiple choice questions and 3 fill-in-the-blank questions correctly, while Jason answered 6 multiple choice questions and 5 fill-in-the-blank questions correctly. If Julia got a total of 14 points and Jason got 16 points, how many points is each type of question worth? Could you use a graph to solve the system of linear equations representing this scenario. In this Concept, you'll learn to solve linear systems by graphing so that you can answer questions like this.

Guidance

In a previous concept, you learned that the intersection of two sets is joined by the word "and." This word also joins two or more equations or inequalities. A set of algebraic sentences joined by the word "and" is called a **system.**

The **solution(s)** to a system is the set of ordered pairs that is in common to each algebraic sentence.

Example A

Determine which of the points (1, 3), (0, 2), *and* (2, 7) *is a solution to the following system of equations.* $\begin{cases} y = 4x - 1 \\ y = 2x + 3 \end{cases}$

Solution: A solution to a system is an ordered pair that is a solution to all the algebraic sentences. To determine if a particular ordered pair is a solution, substitute the coordinates for the variables x and y in each sentence and check.

Check (1, 3) : $\begin{cases} 3 = 4(1) - 1;\ 3 = 3.\ Yes,\ this\ ordered\ pairs\ checks. \\ 3 = 2(1) + 3;\ 3 = 5.\ No,\ this\ ordered\ pair\ does\ not\ check. \end{cases}$

Check (0, 2) : $\begin{cases} 2 = 4(0) - 1;\ 2 = -1.\ No,\ this\ ordered\ pair\ does\ not\ check. \\ 2 = 2(0) + 3;\ 2 = 3.\ No,\ this\ ordered\ pair\ does\ not\ check. \end{cases}$

Check (2, 7) : $\begin{cases} 7 = 4(2) - 1;\ 7 = 7.\ Yes,\ this\ ordered\ pairs\ checks. \\ 7 = 2(2) + 3;\ 7 = 7.\ Yes,\ this\ ordered\ pairs\ checks. \end{cases}$

Because the coordinate (2, 7) works in both equations simultaneously, it is a solution to the system.

To determine the coordinate that is in common to each sentence in the system, each equation can be graphed. The point at which the lines **intersect** represents the solution to the system. The solution can be written two ways:

- As an ordered pair, such as (2, 7)
- By writing the value of each variable, such as $x = 2,\ y = 7$.

Example B

Find the solution to the system.

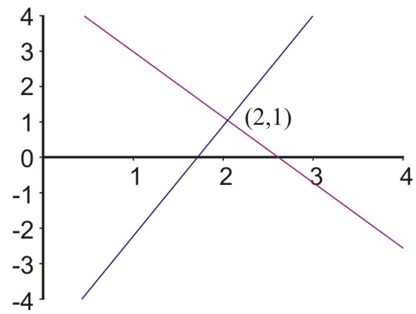
$$\begin{cases} y = 3x - 5 \\ y = -2x + 5 \end{cases}$$

.

Solution: By graphing each equation and finding the point of intersection, you find the solution to the system.

Each equation is written in slope-intercept form and can be graphed using the methods learned previously.

The lines appear to intersect at the ordered pair (2, 1). Is this the solution to the system?



$$\begin{cases} 1 = 3(2) - 5; \quad 1 = 1 \\ 1 = -2(2) + 5; \ 1 = 1 \end{cases}$$

The coordinates check in both sentences. Therefore, (2, 1) is a solution to the system $\begin{cases} y = 3x - 5 \\ y = -2x + 5 \end{cases}$.

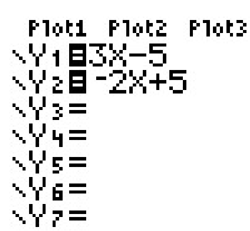
The greatest strength of the graphing method is that it offers a very visual representation of a system of equations and its solution. You can see, however, that determining a solution from a graph would require very careful graphing of the lines and is really practical only when you are certain that the solution gives integer values for x and y. In most cases, this method can offer only approximate solutions to systems of equations. For exact solutions, other methods are necessary.

Solving Systems Using a Graphing Calculator

A graphing calculator can be used to find or check solutions to a system of equations. To solve a system graphically, you must graph the two lines on the same coordinate axes and find the point of intersection. You can use a graphing calculator to graph the lines as an alternative to graphing the equations by hand.

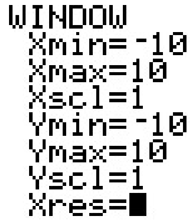
Example C

Using the system from the above example, $\begin{cases} y = 3x - 5 \\ y = -2x + 5 \end{cases}$, we will use a graphing calculator to find the approximate solutions to the system.

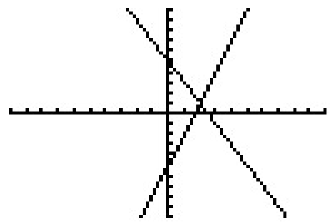


Begin by entering the equations into the $Y =$ menu of the calculator.

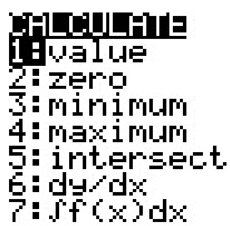
You already know the solution to the system is (2, 1). The window needs to be adjusted so an accurate picture is seen. Change your window to the **default window**.



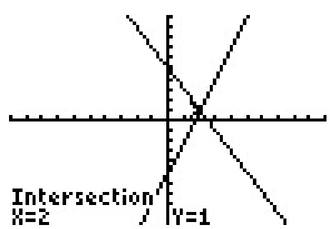
See the graphs by pressing the **GRAPH** button.



The solution to a system is the intersection of the equations. To find the intersection using a graphing calculator, locate the **Calculate** menu by pressing 2^{nd} and **TRACE**. Choose option #5 – **INTERSECTION**.

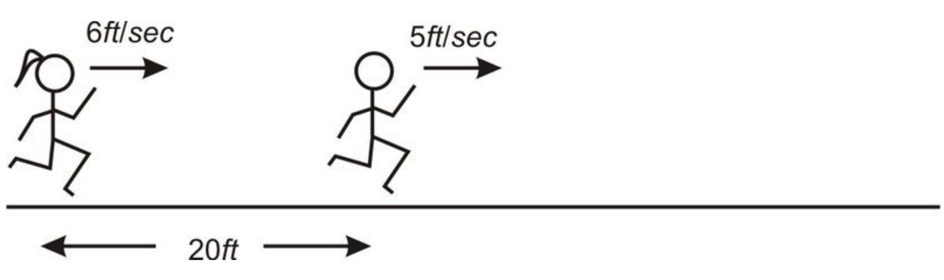


The calculator will ask you **"First Curve?"** Hit **ENTER**. The calculator will automatically jump to the other curve and ask you **"Second Curve?"** Hit **ENTER**. The calculator will ask, **"Guess?"** Hit **ENTER**. The intersection will appear at the bottom of the screen.



Example D

Peter and Nadia like to race each other. Peter can run at a speed of 5 feet per second and Nadia can run at a speed of 6 feet per second. To be a good sport, Nadia likes to give Peter a head start of 20 feet. How long does Nadia take to catch up with Peter? At what distance from the start does Nadia catch up with Peter?

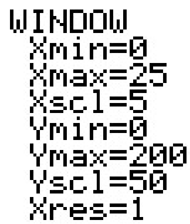


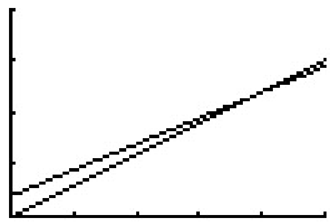
Solution: Begin by translating each runner's situation into an algebraic sentence using $distance = rate \times time$.

Peter: $d = 5t + 20$

Nadia: $d = 6t$

The question asks when Nadia catches Peter. The solution is the point of intersection of the two lines. Graph each equation and find the intersection.





The two lines cross at the coordinate $t = 20$, $d = 120$. This means after 20 seconds Nadia will catch Peter. At this time, they will be at a distance of 120 feet. Any time after 20 seconds Nadia will be farther from the starting line than Peter.

Video Review

MEDIA

Click image to the left for more content.

MEDIA

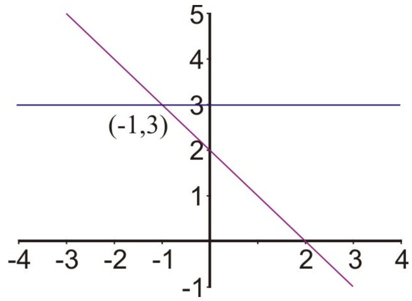
Click image to the left for more content.

Guided Practice

Solve the system $\begin{cases} x + y = 2 \\ \quad\; y = 3 \end{cases}$.

Solution: The first equation is written in standard form. Using its intercepts will be the easiest way to graph this line.

The second equation is a horizontal line three units up from the origin.



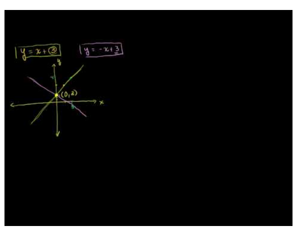
The lines appear to intersect at (–1, 3).

$$\begin{cases} -1+3=2;\ 2=2 \\ 3=3 \end{cases}$$

The coordinates are a solution to each sentence and are a solution to the system.

Practice

Sample explanations for some of the practice exercises below are available by viewing the following video. Note that there is not always a match between the number of the practice exercise in the video and the number of the practice exercise listed in the following exercise set. However, the practice exercise is the same in both. CK-12 Basic Algebra: Solving Linear Systems by Graphing (8:30)



MEDIA

Click image to the left for more content.

1. Define a *system*.
2. What is the solution to a system?
3. Explain the process of solving a system by graphing.
4. What is one problem with using a graph to solve a system?
5. What are the two main ways to write the solution to a system of equations?
6. Suppose Horatio says the solution to a system is (4, –6). What does this mean visually?
7. Where is the **"Intersection"** command located on your graphing calculator? What does it do?
8. In the race example, who is farther from the starting line at 19.99 seconds? At 20.002 seconds?

Determine which ordered pair satisfies the system of linear equations.

9. $\begin{cases} y=3x-2 \\ y=-x \end{cases}$;$(1,4),\ (2,9),\ \left(\frac{1}{2},\ -\frac{1}{2}\right)$
10. $\begin{cases} y=2x-3 \\ y=x+5 \end{cases}$;$(8,13),\ (-7,6),\ (0,4)$
11. $\begin{cases} 2x+y=8 \\ 5x+2y=10 \end{cases}$;$(-9,1),\ (-6,20),\ (14,2)$

12. $\begin{cases} 3x+2y=6 \\ y=\frac{x}{2}-3 \end{cases}$; $\left(3,-\frac{3}{2}\right)$, $(-4,3)$, $\left(\frac{1}{2},4\right)$

In 13 – 22, solve the following systems by graphing.

13.

$$y=x+3$$
$$y=-x+3$$

14.

$$y=3x-6$$
$$y=-x+6$$

15.

$$2x=4$$
$$y=-3$$

16.

$$y=-x+5$$
$$-x+y=1$$

17.

$$x+2y=8$$
$$5x+2y=0$$

18.

$$3x+2y=12$$
$$4x-y=5$$

19.

$$5x+2y=-4$$
$$x-y=2$$

20.

$$2x+4=3y$$
$$x-2y+4=0$$

21.

$$y=\frac{x}{2}-3$$
$$2x-5y=5$$

22.

$$y=4$$
$$x=8-3y$$

23. Mary's car is 10 years old and has a problem. The repair man indicates that it will cost her \$1200.00 to repair her car. She can purchase a different, more efficient car for \$4,500.00. Her present car averages about \$2,000.00 per year for gas while the new car would average about \$1,500.00 per year. Find the number of years for which the total cost of repairs would equal the total cost of replacement.
24. Juan is considering two cell phone plans. The first company charges \$120.00 for the phone and \$30 per month for the calling plan that Juan wants. The second company charges \$40.00 for the same phone, but charges \$45 per month for the calling plan that Juan wants. After how many months would the total cost of the two plans be the same?
25. A tortoise and hare decide to race 30 feet. The hare, being much faster, decided to give the tortoise a head start of 20 feet. The tortoise runs at 0.5 feet/sec and the hare runs at 5.5 feet per second. How long will it be until the hare catches the tortoise?

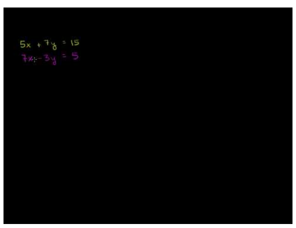
4.2 Solving Systems by Multiplying Both Equations to Cancel a Variable

Here you will solve linear systems using linear combinations in which both equations must be multiplied by a constant to cancel a variable.

A one pound mix consisting of 30% cashews and 70% pistachios sells for \$6.25. A one pound mix consisting of 80% cashews and 20% pistachios sells for \$7.50. How would a mix consisting of 50% of each type of nut sell for?

Watch This

Watch the second portion of this video, starting around 5:00.



MEDIA

Click image to the left for more content.

Khan Academy: Solving Systems of Equations by Multiplication

Guidance

In the linear systems in this lesson, we will need to multiply both equations by a constant in order to have opposite coefficients of one of the variables. In order to determine what numbers to multiply by, we will be finding the least common multiple of the given coefficients. Recall that the least common multiple of two numbers is the smallest number which is divisible by both of the given numbers. For example, 12 is the least common multiple of 4 and 6 because it is the smallest number divisible by both 4 and 6.

Example A

Solve the system using linear combinations:

$$\begin{aligned}2x-5y&=15\\3x+7y&=8\end{aligned}$$

Solution: In this problem we cannot simply multiply one equation by a constant to get opposite coefficients for one of the variables. Here we will need to identify the least common multiple of the coefficients of one of the variables and use this value to determine what to multiple each equation by. If we look at the coefficients of x, 2 and 3, the least common multiple of these numbers is 6. So, we want to have the coefficients of x be 6 and -6 so that they are opposites and will cancel when we add the two equations together. In order to get coefficients of 6 and -6 we can multiply the first equation by 3 and the second equation by -2 (it doesn't matter which one we make negative.)

$$\begin{aligned} 3(2x-5y=15) \quad &\Rightarrow \quad \cancel{6x}-15y=45 \\ -2(3x+7y=8) \quad & \quad + \underline{-\cancel{6x}-14y=-16} \\ & \qquad -29y=29 \\ & \qquad y=-1 \end{aligned}$$

Now find x:

$$\begin{aligned} 2(x)-5(-1)&=15 \\ 2x+5&=15 \\ 2x&=10 \\ x&=5 \end{aligned}$$

Solution: (5, -1)

Check your answer:

$$\begin{aligned} 2(5)-5(-1)&=10+5=15 \\ 3(5)+7(-1)&=15-7=8 \end{aligned}$$

* This problem could also be solved by eliminating the y variables first. To do this, find the least common multiple of the coefficients of y, 5 and 7. The least common multiple is 35, so we would multiply the first equation by 7 and the second equation by 5. Since one of them is already negative, we don't have to multiply by a negative.

Example B

Solve the system using linear combinations:

$$\begin{aligned} 7x+20y&=-9 \\ -2x-3y&=8 \end{aligned}$$

Solution: The first step is to decide which variable to eliminate. Either one can be eliminated but sometimes it is helpful to look at what we need to multiply by to eliminate each one and determine which is easier to eliminate. In general, it is easier to work with smaller numbers so in this case it makes sense to eliminate x first. The Least Common Multiple (LCM) of 7 and 2 is 14. To get 14 as the coefficient of each term, we need to multiple the first equation by 2 and the second equation by 7:

$$\begin{aligned} 2(7x+20y=-9) \quad &\Rightarrow \quad \cancel{14x}+40y=-18 \\ 7(-2x-3y=8) \quad & \quad \underline{-\cancel{14x}-21y=56} \\ & \qquad 19y=38 \\ & \qquad y=2 \end{aligned}$$

Now find x:

$$\begin{aligned}-2x-3(2)&=8\\-2x-6&=8\\-2x&=14\\x&=-7\end{aligned}$$

Solution: (-7, 2)

Check your answer:

$$\begin{aligned}7(-7)+20(2)&=-49+40=-9\\-2(-7)-3(2)&=14-6=8\end{aligned}$$

Example C

Solve the system using linear combinations:

$$\begin{aligned}14x-6y&=-3\\16x-9y&=-7\end{aligned}$$

Solution: This time, we will eliminate y. We need to find the LCM of 6 and 9. The LCM is 18, so we will multiply the first equation by 3 and the second equation by -2. Again, it doesn't matter which equation we multiply by a negative value.

$$\begin{aligned}3(14x-6y=-3) \quad &\Rightarrow \quad +42x-\cancel{18y}=-9\\-2(16x-9y=-7) \quad & \quad \underline{-32x+\cancel{18y}=14}\\ & \qquad\qquad 10x=5\\ & \qquad\qquad x=2\end{aligned}$$

Now find y:

$$\begin{aligned}14\left(\frac{1}{2}\right)-6y&=-3\\7-6y&=-3\\-6y&=-10\\y&=\frac{10}{6}=\frac{5}{3}\end{aligned}$$

Solution: $\left(\frac{1}{2},\frac{5}{3}\right)$

Check your answer:

$$\begin{aligned}14\left(\frac{1}{2}\right)-6\left(\frac{5}{3}\right)&=7-10=-3\\16\left(\frac{1}{2}\right)-9\left(\frac{5}{3}\right)&=8-15=-7\end{aligned}$$

Intro Problem Revisit Write a system of linear equations to represent the given information. Let $x =$ the cost of the cashews per pound and let $y =$ the cost of the pistachios per pound. Now we can write two equations to represent the two different mixes of nuts:

$$\begin{aligned} 0.3x+0.7y &= 6.25 \\ 0.8x+0.2y &= 7.50 \end{aligned}$$

Solve this system to determine the cost of each type of nut per pound. If we eliminate y, we will need to multiple the first equation by 2 and the second equation by -7:

$$\begin{aligned} 2(0.3x-0.7y=6.25) \quad &\Rightarrow \quad +0.6x-\cancel{1.4y}=12.5 \\ -7(0.8x+0.2y=7.50) \quad & \qquad \underline{-5.6x+\cancel{1.4y}=-52.5} \\ & \qquad\qquad -5x=-40 \\ & \qquad\qquad\quad x=8 \end{aligned}$$

Now find y:

$$\begin{aligned} 0.3(8)+0.7y &= 6.25 \\ 2.4+0.7y &= 6.25 \\ 0.7y &= 3.85 \\ y &= 5.5 \end{aligned}$$

So, we have determined that the cost of the cashews is \$8 per pound and the cost of the pistachios is \$5.50 per pound. Now we can determine the cost of the 50% mix as follows:

$0.5(8.00)+0.5(5.50) = 4.00+2.75 = 6.75$ So, the new mix is \$6.75 per pound.

Guided Practice

Solve the following systems using linear combinations:

1.

$$\begin{aligned} 6x+5y &= 3 \\ -4x-2y &= -14 \end{aligned}$$

2.

$$\begin{aligned} 9x-7y &= -19 \\ 5x-3y &= -15 \end{aligned}$$

3.

$$\begin{aligned} 15x-21y &= -63 \\ 7y &= 5x+21 \end{aligned}$$

Answers

1. We can eliminate either variable here. To eliminate x, we can multiple the first equations by 2 and the second equation by 3 to get 12 - the LCM of 6 and 4.

$$\begin{array}{rcr} 2(6x+5y=3) & \Rightarrow & +\cancel{12x}+10y=6 \\ 3(-4x-2y=-14) & & \underline{-\cancel{12x}-6y=-42} \\ & & 4y=-36 \\ & & y=-9 \end{array}$$

Now find x:

$$\begin{aligned} 6x+5(-9)&=3 \\ 6x-45&=3 \\ 6x&=48 \\ x&=8 \end{aligned}$$

Solution: (8, -9)

2. Again we can eliminate either variable. To eliminate y, we can multiply the first equation by 3 and the second equation by -7:

$$\begin{array}{rcr} 3(9x-7y=-19) & \Rightarrow & +27x-\cancel{21y}=-57 \\ -7(5x-3y=-15) & & +\ \underline{-35x+\cancel{21y}=105} \\ & & -8x=48 \\ & & x=-6 \end{array}$$

Now find y:

$$\begin{aligned} 5(-6)-3y&=-15 \\ -30-3y&=-15 \\ -3y&=15 \\ y&=-5 \end{aligned}$$

Solution: (-6, -5)

3. To start this one we need to get the second equation in standard form. The resulting system will be:

$$\begin{aligned} 15x-21y&=-63 \\ -5x+7y&=21 \end{aligned}$$

This time we just need to multiply the second equation by 3 to eliminate x:

$$\begin{aligned} 15x - 21y &= -63 & \Rightarrow \quad & +\cancel{15x} - 21y = -63 \\ 3(-5x + 7y &= 21) & + \quad & \underline{-\cancel{15x} + 21y = -63} \\ & & & 0y = 0 \\ & & & 0 = 0 \end{aligned}$$

Solution: There are infinite solutions.

Practice

Solve the systems using linear combinations.

1.

$$\begin{aligned} 17x - 5y &= 4 \\ 2x + 7y &= 46 \end{aligned}$$

2.

$$\begin{aligned} 9x + 2y &= -13 \\ 11x + 5y &= 2 \end{aligned}$$

3.

$$\begin{aligned} 3x + 4y &= -16 \\ 5x + 5y &= -5 \end{aligned}$$

4.

$$\begin{aligned} 5x - 10y &= 60 \\ 6x + 3y &= -33 \end{aligned}$$

5.

$$\begin{aligned} 3x + 10y &= -50 \\ -5x - 7y &= 6 \end{aligned}$$

6.

$$\begin{aligned} 11x + 6y &= 30 \\ 13x - 5y &= -25 \end{aligned}$$

7.

$$\begin{aligned} 15x + 2y &= 23 \\ 18x - 9y &= -18 \end{aligned}$$

8.

$$
\begin{aligned}
12x + 8y &= 64 \\
17x - 12y &= 9
\end{aligned}
$$

9.

$$
\begin{aligned}
11x - 3y &= 12 \\
33x - 36 &= 9y
\end{aligned}
$$

10.

$$
\begin{aligned}
4x + 3y &= 0 \\
6x - 13y &= 35
\end{aligned}
$$

11.

$$
\begin{aligned}
18x + 2y &= -2 \\
-12x - 3y &= -7
\end{aligned}
$$

12.

$$
\begin{aligned}
-6x + 11y &= -109 \\
8x - 15y &= 149
\end{aligned}
$$

13.

$$
\begin{aligned}
8x &= -5y - 1 \\
-32x + 20y &= 8
\end{aligned}
$$

14.

$$
\begin{aligned}
10x - 16y &= -12 \\
-15x + 14y &= -27
\end{aligned}
$$

Set up and solve a system of equations to answer the following questions.

15. A mix of 35% almonds and 65% peanuts sells for \$5.70. A mix of 75% almonds and 25% peanuts sells for \$6.50. How much should a mix of 60% almonds and 40% peanuts sell for?

16. The Robinson family pays \$19.75 at the movie theater for 3 medium popcorns and 4 medium drinks. The Jamison family pays \$33.50 at the same theater for 5 medium popcorns and 7 medium drinks. How much would it cost for a couple to get 2 medium drinks and 2 medium popcorns?

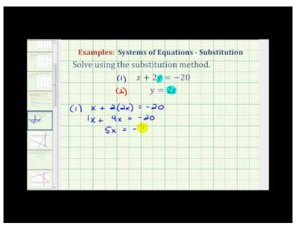
17. A cell phone company charges extra when users exceed their included call time and text message limits. One user paid \$3.24 extra having talked for 240 extra minutes and sending 12 additional texts. A second user talked for 120 extra minutes and sent 150 additional texts and was charged \$4.50 above the regular fee. How much extra would a user be charged for talking 140 extra minutes and sending 200 additional texts?

4.3 Solving Systems with One Solution Using Substitution

Here you will solve consistent, independent systems using the substitution method.

Rex and Carl are making a mixture in science class. They need to have 12 ounces of a 60% saline solution. To make this solution they have a 20% saline solution and an 80% saline solution. How many ounces of each do they need to make the correct mixture?

Watch This



MEDIA

Click image to the left for more content.

James Sousa: Ex 1: Solve a System of Equations Using Substitution

Guidance

In the substitution method we will be looking at the two equations and deciding which variable is easiest to solve for so that we can write one of the equations as $x =$ or $y =$. Next we will replace either the x or the y accordingly in the *other* equation. The result will be an equation with only one variable that we can solve

Example A

Solve the system using substitution:

$$\begin{aligned} 2x+y&=12 \\ -3x-5y&=-11 \end{aligned}$$

Solution: The first step is to look for a variable that is easy to isolate. In other words, does one of the variables have a coefficient of 1? Yes, that variable is the y in the first equation. So, start by isolating or solving for y: $y=-2x+12$

This expression can be used to replace the y in the other equation and solve for x:

$$\begin{aligned} -3x-5(-2x+12)&=-11 \\ -3x+10x-60&=-11 \\ 7x-60&=-11 \\ 7x&=49 \\ x&=7 \end{aligned}$$

Now that we have found x, we can use this value in our expression to find y:

$$\begin{aligned}
y &= -2(7)+12 \\
y &= -14+12 \\
y &= -2
\end{aligned}$$

Recall that the solution to a linear system is a point in the coordinate plane where the two lines intersect. Therefore, our answer should be written as a point: (7, -2). You can check your answer by substituting this point into both equations to make sure that it satisfies them:

$$\begin{aligned}
2(7)+-2 &= 14-2 = 12 \\
-3(7)-5(-2) &= -21+10 = -11
\end{aligned}$$

Example B

Solve the system using substitution:

$$\begin{aligned}
2x+3y &= 13 \\
x+5y &= -4
\end{aligned}$$

Solution: In the last example, y was the easiest variable to isolate. Is that the case here? No, this time, x is the variable with a coefficient of 1. It is easy to fall into the habit of always isolating y since you have done it so much to write equation in slope-intercept form. Try to avoid this and look at each system to see which variable is easiest to isolate. Doing so will help reduce your work.

Solving the second equation for x gives: $x = -5y-4$.

This expression can be used to replace the x in the other equation and solve for y:

$$\begin{aligned}
2(-5y-4)+3y &= 13 \\
-10y-8+3y &= 13 \\
-7y-8 &= 13 \\
-7y &= 21 \\
y &= -3
\end{aligned}$$

Now that we have found y, we can use this value in our expression to find x:

$$\begin{aligned}
x &= -5(-3)-4 \\
x &= 15-4 \\
x &= 11
\end{aligned}$$

So, the solution to this system is (11, -3). Don't forget to check your answer:

$$\begin{aligned}
2(11)+3(-3) &= 22-9 = 13 \\
11+5(-3) &= 11-15 = -4
\end{aligned}$$

Example C

Solve the system using substitution:

$$\begin{aligned}4x+3y&=4\\6x-2y&=19\end{aligned}$$

Solution: In this case, none of the variables have a coefficient of 1. So, we can just pick on to solve for. Let's solve for the x in equation 1:

$$\begin{aligned}4x&=-3y+4\\x&=-\frac{3}{4}y+1\end{aligned}$$

Now, this expression can be used to replace the x in the other equation and solve for y:

$$\begin{aligned}6\left(-\frac{3}{4}y+1\right)-2y&=19\\-\frac{18}{4}y+6-2y&=19\\-\frac{9}{2}y-\frac{4}{2}y&=13\\-\frac{13}{2}y&=13\\\left(-\frac{2}{13}\right)\left(-\frac{13}{2}\right)y&=13\left(-\frac{2}{13}\right)\\y&=-2\end{aligned}$$

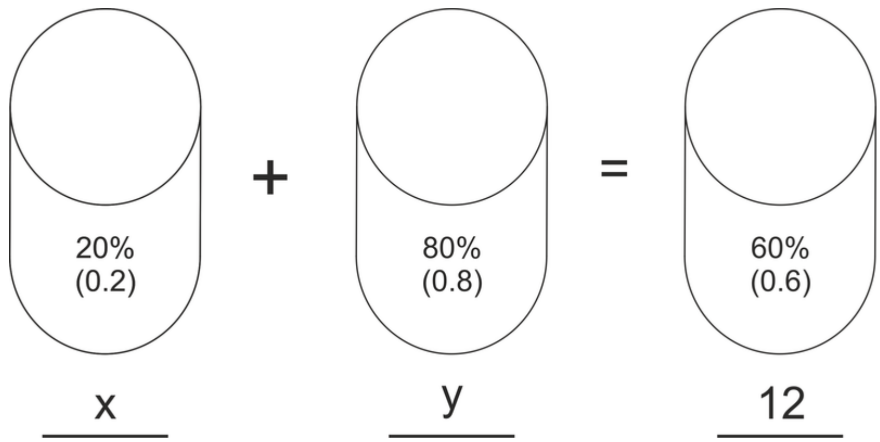
Now that we have found y, we can use this value in our expression find x:

$$\begin{aligned}x&=\left(-\frac{3}{4}\right)(-2)+1\\x&=\frac{6}{4}+1\\x&=\frac{3}{2}+\frac{2}{2}\\x&=\frac{5}{2}\end{aligned}$$

So, the solution is $\left(\frac{5}{2},-2\right)$. Check your answer:

$$\begin{aligned}4\left(\frac{5}{2}\right)+3(-2)&=10-6=4\\6\left(\frac{5}{2}\right)-2(-2)&=15+4=19\end{aligned}$$

Intro Problem Revisit Let's try to make the word problem easier by organizing our information into a "picture" equation as shown below:



In this picture, we can see that we will be mixing x ounces of the 20% solution with y ounces of the 80% solution to get 12 ounces of the 60% solution. The two equations are thus:

$$\begin{aligned}0.2x+0.8y &= 0.6(12)\\ x+y &= 12\end{aligned}$$

Now we can solve the system using substitution. Solve for y in the second equation to get:$y = 12-x$.

Now, substitute and solve in the first equation:

$$\begin{aligned}0.2x+0.8(12-x) &= 0.6(12)\\ 0.2x+9.6-0.8x &= 7.2\\ -0.6x &= -2.4\\ x &= 4\end{aligned}$$

Now we can find y:

$$\begin{aligned}y &= 12-x\\ y &= 12-4\\ y &= 8\end{aligned}$$

Therefore, Rex and Carl need 4 ounces of the 20% saline solution and 8 ounces of the 80% saline solution to make the correct mixture.

Guided Practice

Solve the following systems using the substitution method.

1.

$$\begin{aligned}3x+4y &= -13\\ x &= -2y-9\end{aligned}$$

2.

$$-2x-5y=-39$$
$$x+3y=24$$

3.

$$y=\frac{1}{2}x-21$$
$$y=-2x+9$$

Answers

1. In this problem, the second equation is already solved for x so we can use that in the first equation to find y:

$$\begin{aligned}3(-2y-9)+4y&=-13\\-6y-27+4y&=-13\\-2y-27&=-13\\-2y&=14\\y&=-7\end{aligned}$$

Now we can find x:

$$\begin{aligned}x&=-2(-7)-9\\x&=14-9\\x&=5\end{aligned}$$

Therefore the solution is (5, -7).

2. This time the x in the second equation is the easiest variable to isolate: $x=-3y+24$. Let's use this in the first expression to find y:

$$\begin{aligned}-2(-3y+24)-5y&=-39\\6y-48-5y&=-39\\y-48&=-39\\y&=9\end{aligned}$$

Now we can find x:

$$\begin{aligned}x&=-3(9)+24\\x&=-27+24\\x&=-3\end{aligned}$$

Therefore the solution is (-3, 9).

3. In this case, both equations are equal to y. Since $y = y$, by the Reflexive Property of Equality, we can let the right hand sides of the equations be equal too. This is still a substitution problem; it just looks a little different.

$$\begin{aligned}
\frac{1}{2}x - 21 &= -2x + 9\\
2\left(\frac{1}{2}x - 21\right. &= \left.-2x + 9\right)\\
x - 42 &= -4x + 18\\
5x &= 60\\
x &= 12
\end{aligned}$$

Now we can find y:

$$\begin{aligned}
y &= \frac{1}{2}(12) - 21 & & & y &= -2(12) + 9\\
y &= 6 - 21 & & or & y &= -24 + 9\\
y &= -15 & & & y &= -15
\end{aligned}$$

Therefore our solution is (12, -15).

Practice

Solve the following systems using substitution. Remember to check your answers.

1.

$$\begin{aligned}
x + 3y &= -1\\
2x + 9y &= 7
\end{aligned}$$

2.

$$\begin{aligned}
7x + y &= 6\\
x - 2y &= -12
\end{aligned}$$

3.

$$\begin{aligned}
5x + 2y &= 0\\
y &= x - 7
\end{aligned}$$

4.

$$\begin{aligned} 2x - 5y &= 21 \\ x &= -6y + 2 \end{aligned}$$

5.

$$\begin{aligned} y &= x + 3 \\ y &= 2x - 1 \end{aligned}$$

6.

$$\begin{aligned} x + 6y &= 1 \\ -2x - 11y &= -4 \end{aligned}$$

7.

$$\begin{aligned} 2x + y &= 18 \\ -3x + 11y &= -27 \end{aligned}$$

8.

$$\begin{aligned} 2x + 3y &= 5 \\ 5x + 7y &= 8 \end{aligned}$$

9.

$$\begin{aligned} -7x + 2y &= 9 \\ 5x - 3y &= 3 \end{aligned}$$

10.

$$\begin{aligned} 2x - 6y &= -16 \\ -6x + 10y &= 8 \end{aligned}$$

11.

$$\begin{aligned} 2x - 3y &= -3 \\ 8x + 6y &= 12 \end{aligned}$$

12.

$$\begin{aligned} 5x + y &= -3 \\ y &= 15x + 9 \end{aligned}$$

Set up and solve a system of linear equations to answer each of the following word problems.

13. Alicia and Sarah are at the supermarket. Alicia wants to get peanuts from the bulk food bins and Sarah wants to get almonds. The almonds cost \$6.50 per pound and the peanuts cost \$3.50 per pound. Together they buy 1.5 pounds of nuts. If the total cost is \$6.75, how much did each girl get? Set up a system to solve using substitution.
14. Marcus goes to the department store to buy some new clothes. He sees a sale on t-shirts (\$5.25) and shorts (\$7.50). Marcus buys seven items and his total, before sales tax, is \$43.50. How many of each item did he buy?
15. Jillian is selling tickets for the school play. Student tickets are \$3 and adult tickets are \$5. If 830 people buy tickets and the total revenue is \$3104, how many students attended the play?

4.4 Systems of Linear Inequalities

Here you'll learn how to graph and solve a system of two or more linear inequalities. You'll also determine if such systems are consistent or inconsistent.

What if you were given a system of linear inequalities like $6x-2y \geq 3$ and $2y-3x \leq 7$? How could you determine its solution? After completing this Concept, you'll be able to find the solution region of systems of linear inequalities like this one.

Watch This

MEDIA

Click image to the left for more content.

CK-12 Foundation: 0710S Systems of Linear Inequalities

Guidance

In the last chapter you learned how to graph a linear inequality in two variables. To do that, you graphed the equation of the straight line on the coordinate plane. The line was solid for $\leq$ or $\geq$ signs (where the equals sign is included), and the line was dashed for <or >signs (where the equals sign is not included). Then you shaded above the line (if the inequality began with $y >$ or $y \geq$) or below the line (if it began with $y <$ or $y \leq$).

In this section, we'll see how to graph two or more linear inequalities on the same coordinate plane. The inequalities are graphed separately on the same graph, and the solution for the system is the common shaded region between all the inequalities in the system. One linear inequality in two variables divides the plane into two **half-planes**. A **system** of two or more linear inequalities can divide the plane into more complex shapes.

Let's start by solving a system of two inequalities.

Graph a System of Two Linear Inequalities

Example A

Solve the following system:

$$\begin{aligned} 2x+3y &\leq 18 \\ x-4y &\leq 12 \end{aligned}$$

Solution

Solving systems of linear inequalities means graphing and finding the intersections. So we graph each inequality, and then find the intersection *regions* of the solution.

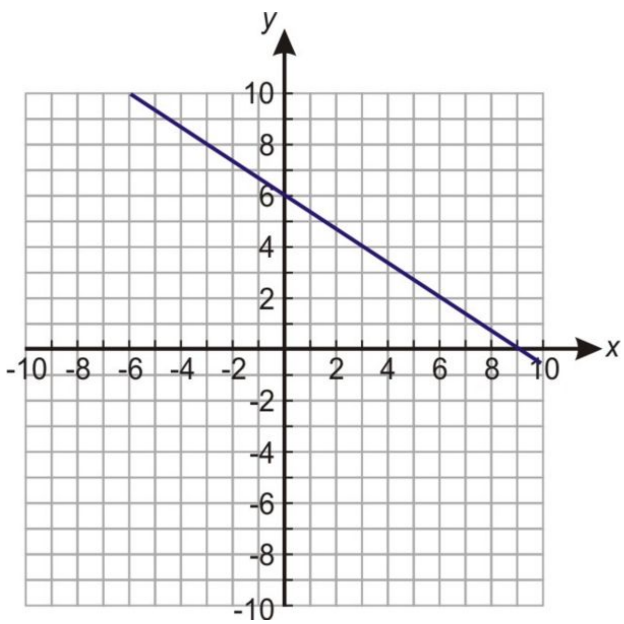
First, let's rewrite each equation in slope-intercept form. (Remember that this form makes it easier to tell which region of the coordinate plane to shade.) Our system becomes

$$\begin{aligned} 3y &\le -2x+18 & \qquad y &\le -\frac{2}{3}x+6 \\ & & \Rightarrow & \\ -4y &\le -x+12 & \qquad y &\ge \frac{x}{4}-3 \end{aligned}$$

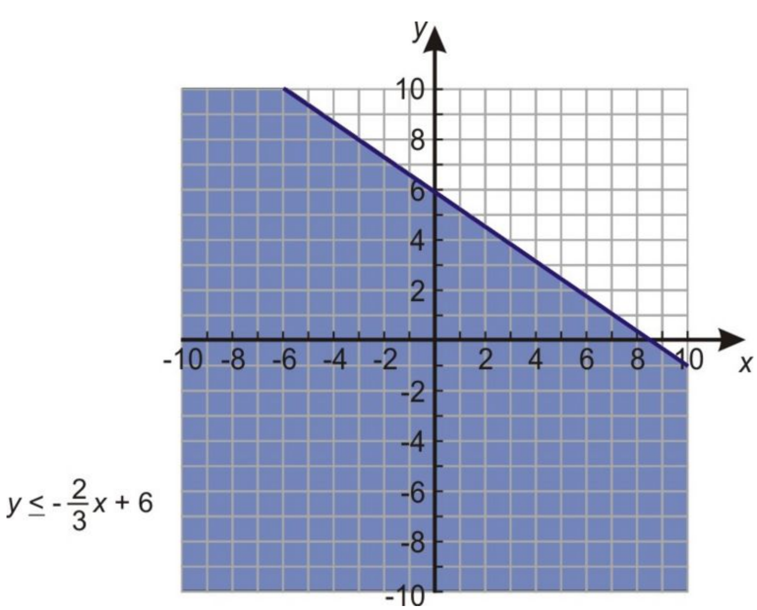
Notice that the inequality sign in the second equation changed because we divided by a negative number!

For this first example, we'll graph each inequality separately and then combine the results.

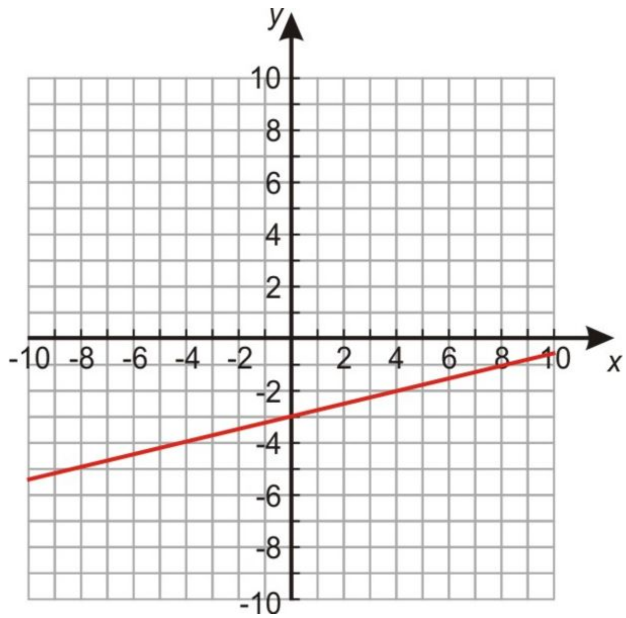
Here's the graph of the first inequality:



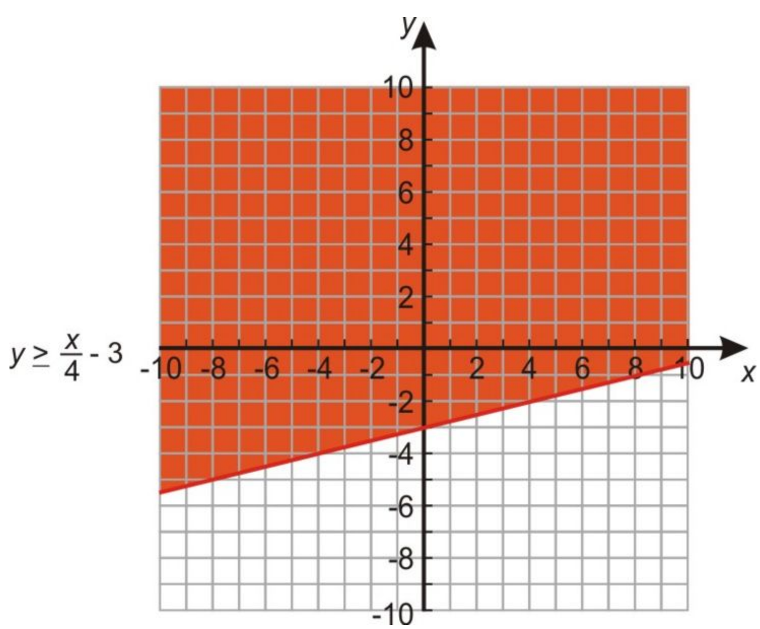
The line is solid because the equals sign is included in the inequality. Since the inequality is **less** than or equal to, we shade **below** the line.



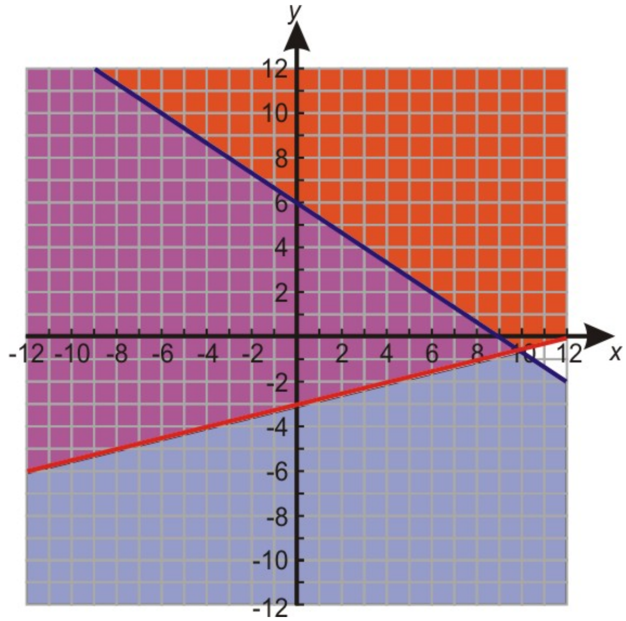
And here's the graph of the second inequality:



The line is solid again because the equals sign is included in the inequality. We now shade **above** the line because y is **greater** than or equal to.



When we combine the graphs, we see that the blue and red shaded regions overlap. The area where they overlap is the area where both inequalities are true. Thus that area (shown below in purple) is the solution of the system.



The kind of solution displayed in this example is called **unbounded**, because it continues forever in at least one direction (in this case, forever upward and to the left).

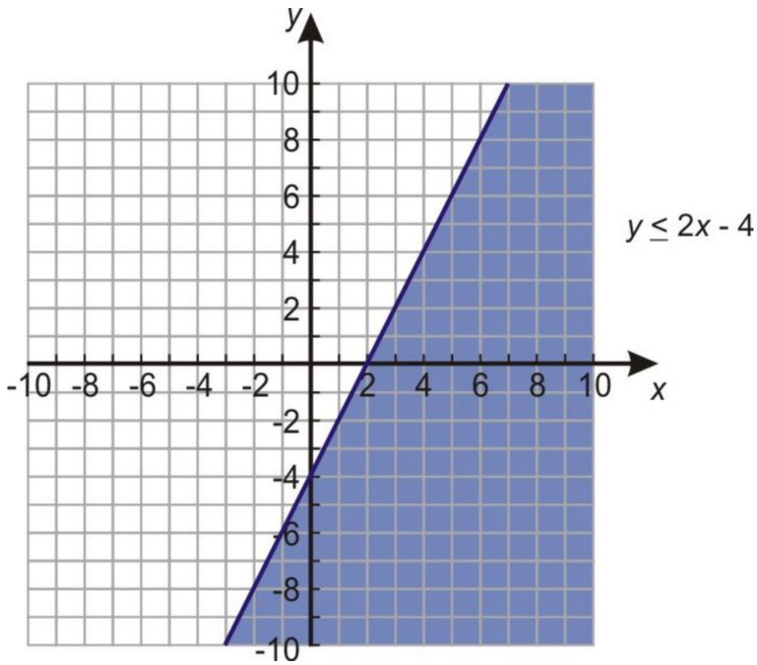
Example B

There are also situations where a system of inequalities has no solution. For example, let's solve this system.

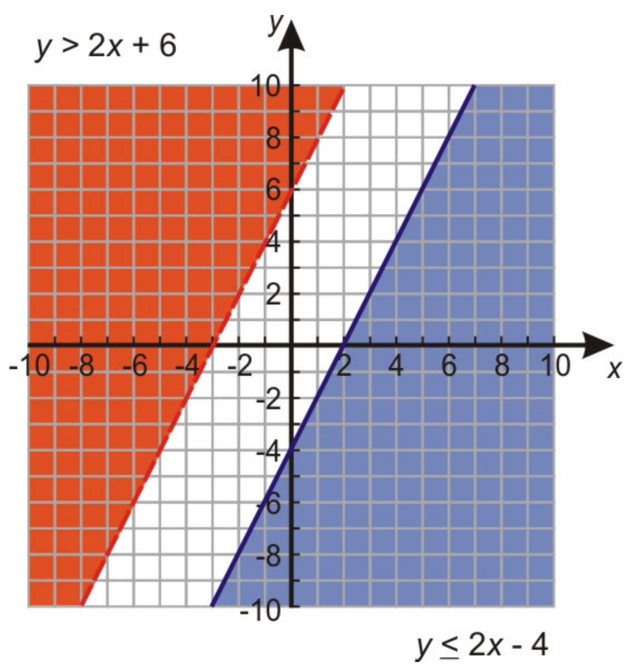
$$y \leq 2x - 4$$
$$y > 2x + 6$$

Solution

We start by graphing the first line. The line will be solid because the equals sign is included in the inequality. We must shade downwards because y is less than.



Next we graph the second line on the same coordinate axis. This line will be dashed because the equals sign is not included in the inequality. We must shade upward because y is greater than.



It doesn't look like the two shaded regions overlap at all. The two lines have the same slope, so we know they are parallel; that means that the regions indeed won't ever overlap since the lines won't ever cross. So this system of inequalities has no solution.

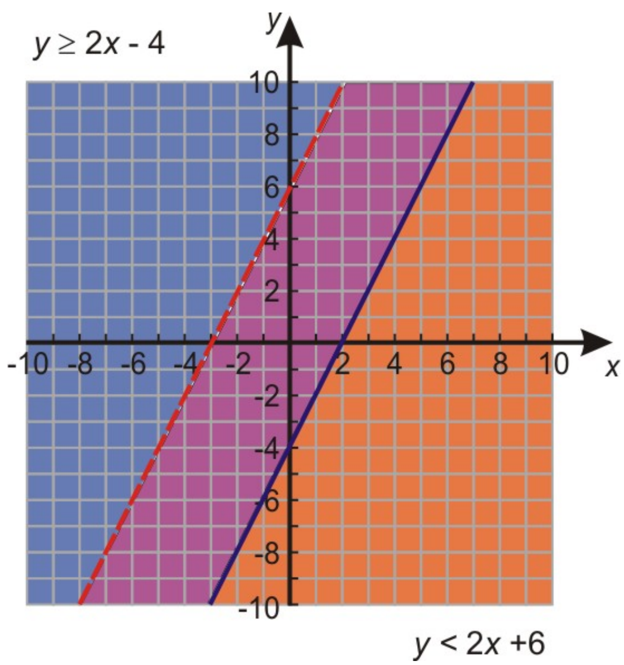
But a system of inequalities can sometimes have a solution even if the lines are parallel. For example, what happens if we swap the directions of the inequality signs in the system we just graphed?

To graph the system

$$y \geq 2x - 4$$
$$y < 2x + 6$$

,

we draw the same lines we drew for the previous system, but we shade *upward* for the first inequality and *downward* for the second inequality. Here is the result:



You can see that this time the shaded regions overlap. The area between the two lines is the solution to the system.

Graph a System of More Than Two Linear Inequalities

When we solve a system of just two linear inequalities, the solution is always an **unbounded** region—one that continues infinitely in at least one direction. But if we put together a system of more than two inequalities, sometimes we can get a solution that is **bounded**—a finite region with three or more sides.

Let's look at a simple example.

Example C

Find the solution to the following system of inequalities.

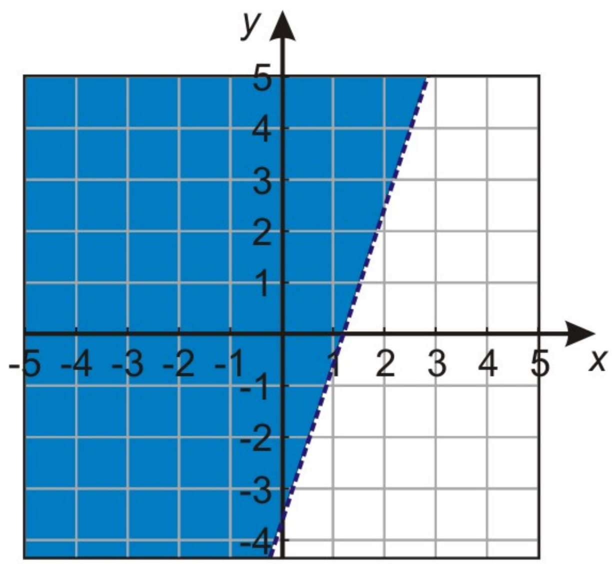
$$3x - y < 4$$
$$4y + 9x < 8$$
$$x \geq 0$$
$$y \geq 0$$

Solution

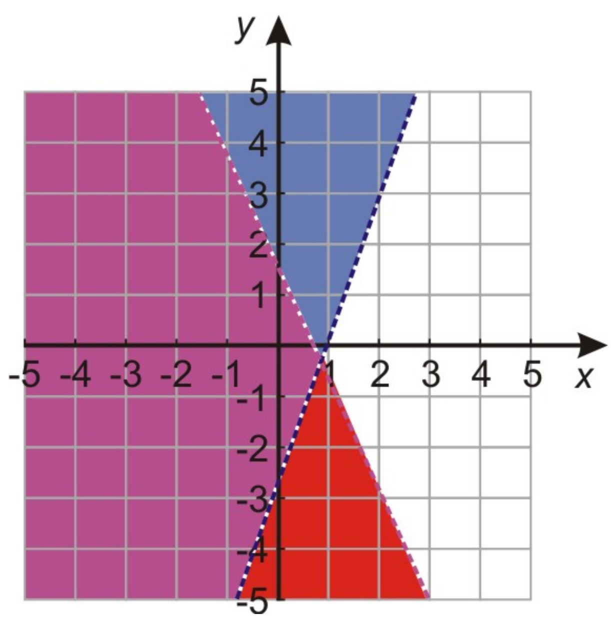
Let's start by writing our inequalities in slope-intercept form.

$$\begin{aligned}
&y > 3x-4\\
&y < -\frac{9}{4}x+2\\
&x \geq 0\\
&y \geq 0
\end{aligned}$$

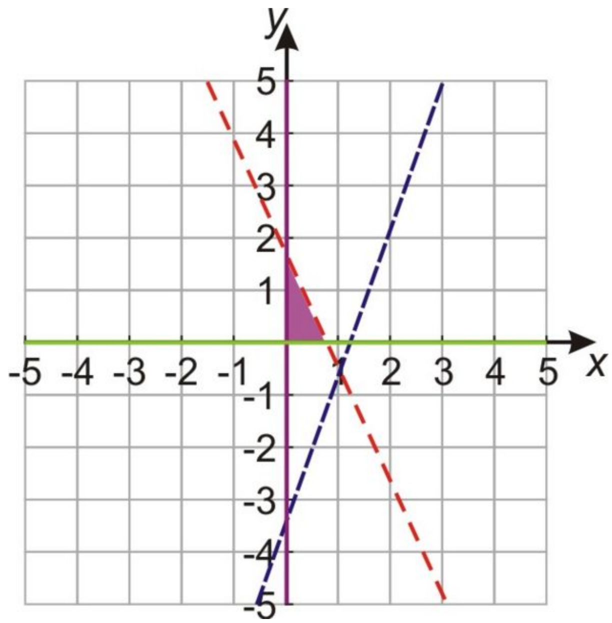
Now we can graph each line and shade appropriately. First we graph $y > 3x-4$:



Next we graph $y < -\frac{9}{4}x+2$:



Finally we graph $x \geq 0$ and $y \geq 0$, and we're left with the region below; this is where all four inequalities overlap.



The solution is **bounded** because there are lines on all sides of the solution region. In other words, the solution region is a bounded geometric figure, in this case a triangle.

Notice, too, that only three of the lines we graphed actually form the boundaries of the region. Sometimes when we graph multiple inequalities, it turns out that some of them don't affect the overall solution; in this case, the solution would be the same even if we'd left out the inequality $y > 3x - 4$. That's because the solution region of the system formed by the other three inequalities is completely contained within the solution region of that fourth inequality; in other words, any solution to the other three inequalities is *automatically* a solution to that one too, so adding that inequality doesn't narrow down the solution set at all.

But that wasn't obvious until we actually drew the graph!

Watch this video for help with the Examples above.

MEDIA

Click image to the left for more content.

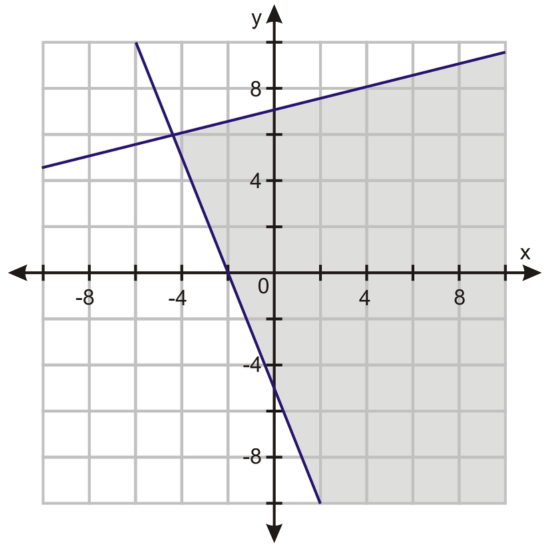
CK-12 Foundation: Systems of Linear Inequalities

Vocabulary

- **Solution for the system of inequalities:** The ***solution for the system of inequalities*** is the common shaded region between all the inequalities in the system.
- **Feasible region:** The common shaded region of the system of inequalities is called the ***feasible region***.
- **Optimization:** The goal is to locate the feasible region of the system and use it to answer a profitability, or ***optimization***, question.

Guided Practice

Write the system of inequalities shown below.



Solution:

There are two boundary lines, so there are two inequalities. Write each one in slope-intercept form.

$$y \leq \frac{1}{4}x + 7$$
$$y \geq -\frac{5}{2}x - 5$$

Practice

1. Consider the system

$$y < 3x - 5$$
$$y > 3x - 5$$

. Is it consistent or inconsistent? Why?
2. Consider the system

$$y \leq 2x + 3$$
$$y \geq 2x + 3$$

. Is it consistent or inconsistent? Why?
3. Consider the system

$$y \leq -x + 1$$
$$y > -x + 1$$

. Is it consistent or inconsistent? Why?
4. In example 3 in this lesson, we solved a system of four inequalities and saw that one of the inequalities, $y > 3x - 4$, didn't affect the solution set of the system.

a. What would happen if we changed that inequality to $y < 3x - 4$?
b. What's another inequality that we could add to the original system without changing it? Show how by sketching a graph of that inequality along with the rest of the system.
c. What's another inequality that we could add to the original system to make it inconsistent? Show how by sketching a graph of that inequality along with the rest of the system.

5. Recall the compound inequalities in one variable that we worked with back in chapter 6. Compound inequalities with "and" are simply systems like the ones we are working with here, except with one variable instead of two.
   a. Graph the inequality $x > 3$ in two dimensions. What's another inequality that could be combined with it to make an inconsistent system?
   b. Graph the inequality $x \leq 4$ on a number line. What two-dimensional system would have a graph that looks just like this one?

Find the solution region of the following systems of inequalities.

6.

$$\begin{aligned} x - y &< -6 \\ 2y &\geq 3x + 17 \end{aligned}$$

7.

$$\begin{aligned} 4y - 5x &< 8 \\ -5x &\geq 16 - 8y \end{aligned}$$

8.

$$\begin{aligned} 5x - y &\geq 5 \\ 2y - x &\geq -10 \end{aligned}$$

9.

$$\begin{aligned} 5x + 2y &\geq -25 \\ 3x - 2y &\leq 17 \\ x - 6y &\geq 27 \end{aligned}$$

10.

$$\begin{aligned} 2x - 3y &\leq 21 \\ x + 4y &\leq 6 \\ 3x + y &\geq -4 \end{aligned}$$

11.

$$\begin{aligned} 12x - 7y &< 120 \\ 7x - 8y &\geq 36 \\ 5x + y &\geq 12 \end{aligned}$$

12. A club is selling t-shirts and pendants at sports events as a fundraiser. Their budget is \$500 and they want to order at least 125 items. They must buy at least as many t-shirts as they buy pendants. Each t-shirt costs \$5 and each pendant costs \$3.

(a) Write a system of inequalities to represent the situation.
(b) Graph the inequalities.
(c) Are there any solutions graphed that do not make sense in the context of the problem? Explain.
(d) Describe the constraints on the inequalities.
(e) If the club buys 50 t-shirts and 100 pendants, will the conditions be satisfied?
(f) What is the maximum number of t-shirts they can buy and still meet the conditions?

Texas Instruments Resources

In the CK-12 Texas Instruments Algebra I FlexBook, there are graphing calculator activities designed to supplement the objectives for some of the lessons in this chapter. See http://www.ck12.org/flexr/chapter/9617.