Math For Our World

MATH FOR OUR WORLD

An open textbook for MATH502

GAIL POITRAST

CONTENTS

Chapter II. Chapter 2: Using Excel

Chapter IV. Chapter 4: Statistics: Collecting Data

Chapter V. Chapter 5: Statistics: Describing Data

Chapter VI. Chapter 6: Statistics: Normal Distribution

Chapter VII. Chapter 7: Probability

INTRODUCTION

Gail Poitrast

This course is based on the text *Math in Society,* edited by David Lippman, Pierce College Ft Steilacoom, and on the text *Using Excel*, by Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker. The modules in this course have been adapted and edited by Gail Poitrast to fit the learning objectives of Math 502.

This course is a survey of contemporary mathematical topics, most non-algebraic, appropriate for a college-level quantitative literacy topics course for liberal arts majors. The text is designed so that most chapters are independent. Emphasis is placed on the applicability of the mathematics. The use of Excel is also explored, as it is a learning objective for Math 502.

This book contains content from the following sources:

COURSE TOPICS

The course's content, assignments, and assessments for Math for Our World are aligned to the following topics:

Chapter 1: General Problem Solving

- Problem Solving with Proportional Relationships and Geometry
- Unit Conversion
- Estimation, Precision and Accuracy

Chapter 2: Excel

- Data Entering, Editing, and Formatting
- Data Analysis, Statistical Functions & Formulas
- Charts: Choosing & Formatting Appropriately

Chapter 3: Math of Finance

- Simple and Compound Interest
- Annuities and Loans

Chapter 4: Statistics: Collecting Data

- Data Collection Basics
- Sampling and Experimentation
- Graphical Representation of Data

Chapter 5: Statistics: Describing Data

- Measure of Central Tendency
- Measures of Distribution

Chapter 6: Normal Distribution

- The Normal Curve
- Z-scores and Percentiles

Chapter 7: Probability

- Computing the Probability of an Event
- Applications With Probability

CHAPTER I

CHAPTER 1: GENERAL PROBLEM SOLVING

WHY IT MATTERS: GENERAL PROBLEM SOLVING

Why understand the basics of problem solving?

Thinking comes naturally. You don't have to make it happen—it just does. But you can make it happen in different ways. For example, you can think positively or negatively. You can think with "heart" and you can think with rational judgment. You can also think strategically and analytically, and mathematically and scientifically. These are a few of multiple ways in which the mind can process thought.

What are some forms of thinking you use? When do you use them, and why?

As a college student, you are tasked with engaging and expanding your thinking skills. One of the most important of these skills is critical thinking. Critical thinking is important because it relates to nearly all tasks, situations, topics, careers, environments, challenges, and opportunities. It's a "domain-general" thinking skill—not a thinking skill that's reserved for a one subject alone or restricted to a particular subject area.

Great leaders have highly attuned critical thinking skills, and you can, too. In fact, you probably have a lot of these skills already. Of all your thinking skills, critical thinking may have the greatest value.

What Is Critical Thinking?

Critical thinking is clear, reasonable, reflective thinking focused on deciding what to believe or do. It means asking probing questions like, "How do we know?" or "Is this true in every case or just in this instance?" It involves being skeptical and challenging assumptions, rather than simply memorizing facts or blindly accepting what you hear or read.

Who are critical thinkers, and what characteristics do they have in common? Critical thinkers are usually curious and reflective people. They like to explore and probe new areas and seek knowledge, clarification, and new solutions. They ask pertinent questions, evaluate statements and arguments, and they distinguish between facts and opinion. They are also willing to examine their own beliefs, possessing a manner of humility that allows them to admit lack of knowledge or understanding when needed. They are open to changing their mind. Perhaps most of all, they actively enjoy learning, and seeking new knowledge is a lifelong pursuit.

This may well be you!

Critical Thinking IS	**Critical Thinking is NOT**
Skepticism	Memorizing
Examining assumptions	Group thinking
Challenging reasoning	Blind acceptance of authority
Uncovering biases	

The following video, from Lawrence Bland, presents the major concepts and benefits of critical thinking.

A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=59

Critical Thinking and Logic

Critical thinking is fundamentally a process of questioning information and data. You may question the information you read in a textbook, or you may question what a politician or a professor or a classmate says. You can also question a commonly-held belief or a new idea. With critical thinking, anything and everything is subject to question and examination for the purpose of logically constructing reasoned perspectives.

Questions of Logic in Critical Thinking

Let's use a simple example of applying logic to a critical-thinking situation. In this hypothetical scenario, a man has a PhD in political science, and he works as a professor at a local college. His wife works at the college, too. They have three young children in the local school system, and their family is well known in the community. The man is now running for political office. Are his credentials and experience sufficient for entering public office? Will he be effective in the political office? Some voters might believe that his personal life and current job, on the surface, suggest he will do well in the position, and they will vote for him. In truth, the characteristics described don't guarantee that the man will do a good job. The information is somewhat irrelevant. What else might you want to know? How about whether the man had already held a political office and done a good job? In this case, we want to ask, How much information is adequate in order to make a decision based on logic instead of assumptions?

The following questions are ones you may apply to formulating a logical, reasoned perspective in the above scenario or any other situation:

1. *What's happening?* Gather the basic information and begin to think of questions.
2. *Why is it important?* Ask yourself why it's significant and whether or not you agree.
3. *What don't I see?* Is there anything important missing?
4. *How do I know?* Ask yourself where the information came from and how it was constructed.
5. *Who is saying it?* What's the position of the speaker and what is influencing them?
6. *What else? What if?* What other ideas exist and are there other possibilities?

Problem-Solving with Critical Thinking

For most people, a typical day is filled with critical thinking and problem-solving challenges. In fact, critical thinking and problem-solving go hand-in-hand. They both refer to using knowledge, facts, and data to solve problems effectively. But with problem-solving, you are specifically identifying, selecting, and defending your solution.

Problem-Solving Action Checklist

Problem-solving can be an efficient and rewarding process, especially if you are organized and mindful of critical steps and strategies. Remember, too, to assume the attributes of a good critical thinker. If you are curious, reflective, knowledge-seeking, open to change, probing, organized, and ethical, your challenge or problem will be less of a hurdle, and you'll be in a good position to find intelligent solutions.

	STRATEGIES	ACTION CHECKLIST[1]
1	Define the problem	• Identify the problem • Provide as many supporting details as possible • Provide examples • Organize the information logically
2	Identify available solutions	• Use logic to identify your most important goals • Identify implications and consequences • Identify facts • Compare and contrast possible solutions
3	Select your solution	• Use gathered facts and relevant evidence • Support and defend solutions considered valid • Defend your solution

1. "Student Success-Thinking Critically In Class and Online." *Critical Thinking Gateway*. St Petersburg College, n.d. Web. 16 Feb 2016.

Critical Thinking, Problem Solving, and Math

In previous math courses, you've no doubt run into the infamous "word problems." Unfortunately, these problems rarely resemble the type of problems we actually encounter in everyday life. In math books, you usually are told exactly which formula or procedure to use, and are given exactly the information you need to answer the question. In real life, problem solving requires identifying an appropriate formula or procedure, and determining what information you will need (and won't need) to answer the question.

In this section, we will review several basic but powerful algebraic ideas: **percents**, **rates**, and **proportions**. We will then focus on the problem solving process, and explore how to use these ideas to solve problems where we don't have perfect information.

PROPORTIONAL RELATIONSHIPS AND A BIT OF GEOMETRY

Learning Outcomes

- Given the part and the whole, write a percent
- Calculate both relative and absolute change of a quantity
- Calculate tax on a purchase

In the 2004 vice-presidential debates, Democratic contender John Edwards claimed that US forces have suffered "90% of the coalition casualties" in Iraq. Incumbent Vice President Dick Cheney disputed this, saying that in fact Iraqi security forces and coalition allies "have taken almost 50 percent" of the casualties.[1]

1. http://www.factcheck.org/cheney_edwards_mangle_facts.html

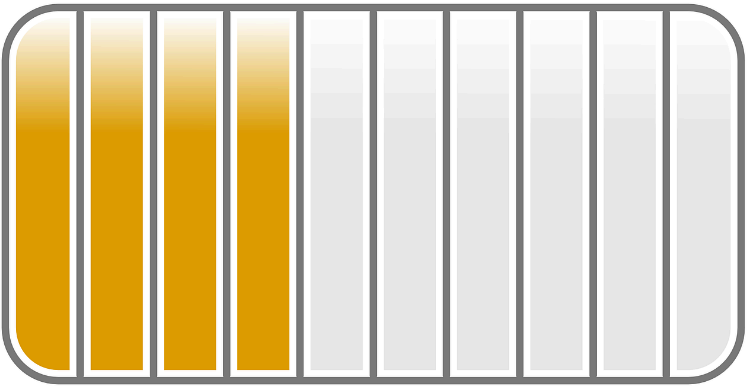
Who was correct? How can we make sense of these numbers?

In this section, we will show how the idea of percent is used to describe parts of a whole. Percents are prevalent in the media we consume regularly, making it imperative that you understand what they mean and where they come from.

We will also show you how to compare different quantities using proportions. Proportions can help us understand how things change or relate to each other.

Percents

Percent literally means "per 100," or "parts per hundred." When we write 40%, this is equivalent to the fraction $\frac{40}{100}$ or the decimal 0.40. Notice that 80 out of 200 and 10 out of 25 are also 40%, since $\frac{80}{200} = \frac{10}{25} = \frac{40}{100}$.



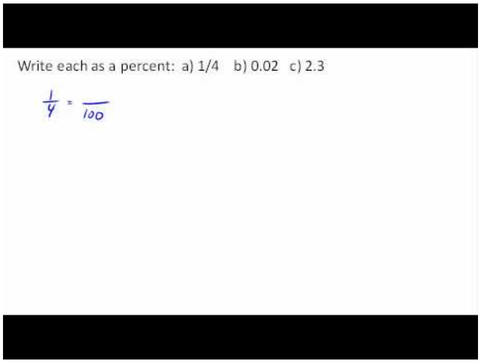
A visual depiction of 40%

Percent

If we have a *part* that is some *percent* of a *whole*, then $\text{percent} = \dfrac{\text{part}}{\text{whole}}$, or equivalently, $\text{part} \cdot \text{whole} = \text{percent}$

To do the calculations, we write the percent as a decimal.

For a refresher on basic percentage rules, using the examples on this page, view the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

Examples

In a survey, 243 out of 400 people state that they like dogs. What percent is this?

Solution:

$\frac{243}{400} = 0.6075 = \frac{60.75}{100}$ This is 60.75%.

Notice that the percent can be found from the equivalent decimal by moving the decimal point two places to the right.

Example

Write each as a percent:

1. $\frac{1}{4}$
2. 0.02
3. 2.35

Solutions:

1. $\frac{1}{4} = 0.25$ = 25%
2. 0.02 = 2%
3. 2.35 = 235%

TRY IT NOW

Throughout this text, you will be given opportunities to answer questions and know immediately whether you answered correctly. To answer the question below, do the calculation on a separate piece of paper and enter your answer in the box. Click on the submit button , and if you are correct, a green box will appear around your answer. If you are incorrect, a red box will appear. You can click on "Try Another Version of This Question" as many times as you like. Practice all you want!

Click here to try!

Example

In the news, you hear "tuition is expected to increase by 7% next year." If tuition this year was $1200 per quarter, what will it be next year?

Solution:

The tuition next year will be the current tuition plus an additional 7%, so it will be 107% of this year's tuition: \$1200(1.07) = \$1284.

Alternatively, we could have first calculated 7% of \$1200: \$1200(0.07) = \$84.

Notice this is not the expected tuition for next year (we could only wish). Instead, this is the expected increase, so to calculate the expected tuition, we'll need to add this change to the previous year's tuition: \$1200 + \$84 = \$1284.

TRY IT NOW

Click here to try!

Example

The value of a car dropped from \$7400 to \$6800 over the last year. What percent decrease is this?

Solution:

To compute the percent change, we first need to find the dollar value change: \$6800 – \$7400 = –\$600. Often we will take the absolute value of this amount, which is called the **absolute change**: |–600| = 600.

Since we are computing the decrease relative to the starting value, we compute this percent out of \$7400:

$\frac{600}{7400} = 0.081 =$ 8.1% decrease. This is called a **relative change**.

Absolute and Relative Change

Given two quantities,

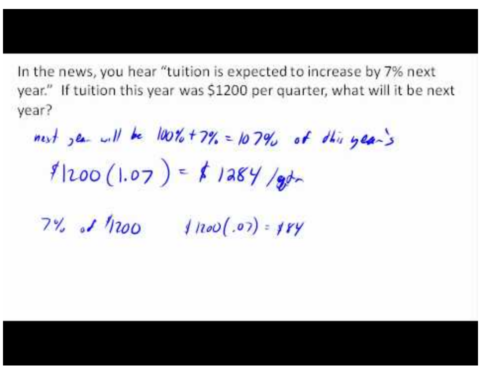
Absolute change = $|\text{ending quantity} - \text{starting quantity}|$

Relative change: $\frac{\text{absolute change}}{\text{starting quantity}}$

- Absolute change has the same units as the original quantity.
- Relative change gives a percent change.

The starting quantity is called the **base** of the percent change.

For a deeper dive on absolute and relative change, using the examples on this page, view the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

The base of a percent is very important. For example, while Nixon was president, it was argued that marijuana was a "gateway" drug, claiming that 80% of marijuana smokers went on to use harder drugs like cocaine. The problem is, this isn't true. The true claim is that 80% of harder drug users first smoked marijuana. The difference is one of base: 80% of marijuana smokers using hard drugs, vs. 80% of hard drug users having smoked marijuana. These numbers are not equivalent. As it turns out, only one in 2,400 marijuana users actually go on to use harder drugs.[2]

Example

There are about 75 QFC supermarkets in the United States. Albertsons has about 215 stores. Compare the size of the two companies.

Solution:

When we make comparisons, we must ask first whether an absolute or relative comparison. The absolute difference is 215 – 75 = 140. From this, we

2. http://tvtropes.org/pmwiki/pmwiki.php/Main/LiesDamnedLiesAndStatistics

could say “Albertsons has 140 more stores than QFC.” However, if you wrote this in an article or paper, that number does not mean much. The relative difference may be more meaningful. There are two different relative changes we could calculate, depending on which store we use as the base:

Using QFC as the base, $\frac{140}{75} = 1.867$.

This tells us Albertsons is 186.7% larger than QFC.

Using Albertsons as the base, $\frac{140}{215} = 0.651$.

This tells us QFC is 65.1% smaller than Albertsons.

Notice both of these are showing percent *differences*. We could also calculate the size of Albertsons relative to QFC: $\frac{215}{75} = 2.867$, which tells us Albertsons is 2.867 times the size of QFC. Likewise, we could calculate the size of QFC relative to Albertsons: $\frac{75}{215} = 0.349$, which tells us that QFC is 34.9% of the size of Albertsons.

Example

Suppose a stock drops in value by 60% one week, then increases in value the next week by 75%. Is the value higher or lower than where it started?

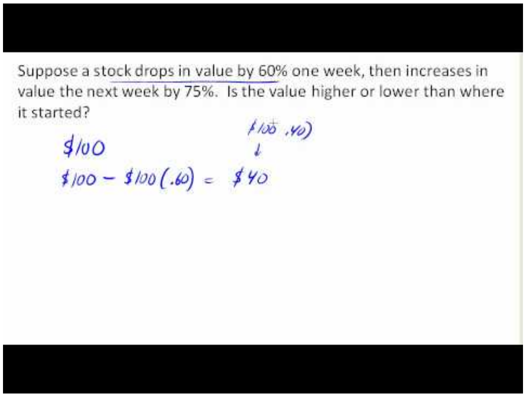
Solution:

To answer this question, suppose the value started at \$100. After one week, the value dropped by 60%: \$100 – \$100(0.60) = \$100 – \$60 = \$40.

In the next week, notice that base of the percent has changed to the new value, \$40. Computing the 75% increase: \$40 + \$40(0.75) = \$40 + \$30 = \$70.

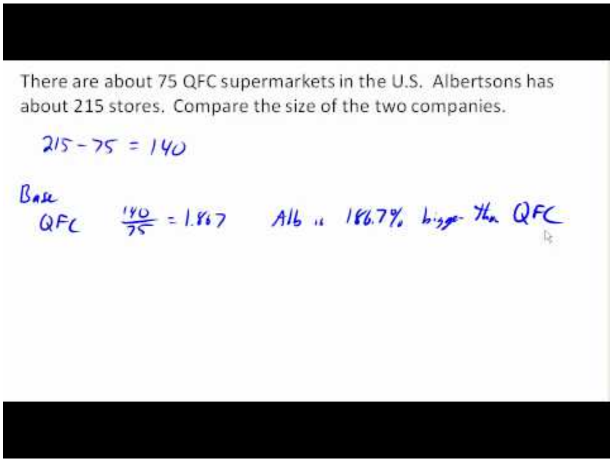
In the end, the stock is still \$30 lower, or $\frac{\$30}{100}$ = 30% lower, valued than it started.

A video walk-through of this example can be seen here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

Consideration of the base of percentages is explored in this video, using the examples on this page.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

TRY IT NOW

[Click here to try!](#)

Example

A *Seattle Times* article on high school graduation rates reported "The number of schools graduating 60 percent or fewer students in four years—sometimes referred to as 'dropout factories'—decreased by 17 during that time period. The number of kids attending schools with such low graduation rates was cut in half."

1. Is the "decreased by 17" number a useful comparison?
2. Considering the last sentence, can we conclude that the number of "dropout factories" was originally 34?

Solutions:

1. This number is hard to evaluate, since we have no basis for judging whether this is a larger or small change. If the number of "dropout factories" dropped from 20 to 3, that'd be a very significant change, but if the number dropped from 217 to 200, that'd be less of an improvement.
2. The last sentence provides relative change,

which helps put the first sentence in perspective. We can estimate that the number of "dropout factories" was probably previously around 34. However, it's possible that students simply moved schools rather than the school improving, so that estimate might not be fully accurate.

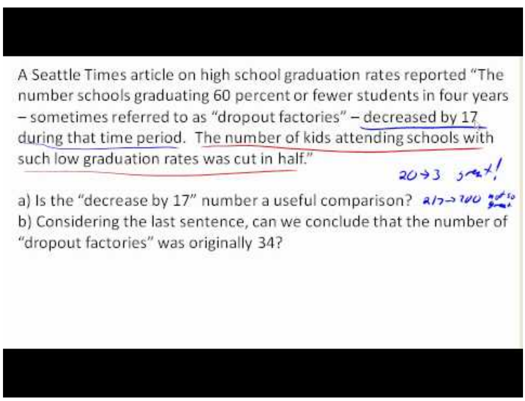
Example

Let's return to the example at the top of this page. In the 2004 vice-presidential debates, Democratic candidate John Edwards claimed that US forces have suffered "90% of the coalition casualties" in Iraq. Cheney disputed this, saying that in fact Iraqi security forces and coalition allies "have taken almost 50 percent" of the casualties. Who is correct?

Solution:

Without more information, it is hard for us to judge who is correct, but we can easily conclude that these two percents are talking about different things, so one does not necessarily contradict the other. Edward's claim was a percent with coalition forces as the base of the percent, while Cheney's claim was a percent with both coalition and Iraqi security forces as the base of the percent. It turns out both statistics are in fact fairly accurate.

A detailed explanation of these examples can be viewed here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

Think About It

In the 2012 presidential elections, one candidate argued that “the president’s plan will cut $716 billion from Medicare, leading to fewer services for seniors,” while the other candidate rebuts that “our plan does not cut current spending and actually expands benefits for seniors, while implementing cost saving measures.” Are these claims in conflict, in agreement, or not comparable because they’re talking about different things?

We’ll wrap up our review of percents with a couple cautions. First, when talking about a change of quantities that are already measured in percents, we have to be careful in how we describe the change.

Example

A politician's support increases from 40% of voters to 50% of voters. Describe the change.

Solution:

We could describe this using an absolute change: $|50\% - 40\%| = 10\%$. Notice that since the original quantities were percents, this change also has the units of percent. In this case, it is best to describe this as an increase of 10 **percentage points**.

In contrast, we could compute the percent change: $\frac{10\%}{40\%} = 0.25 = 25\%$ increase. This is the relative change, and we'd say the politician's support has increased by 25%.

Lastly, a caution against averaging percents.

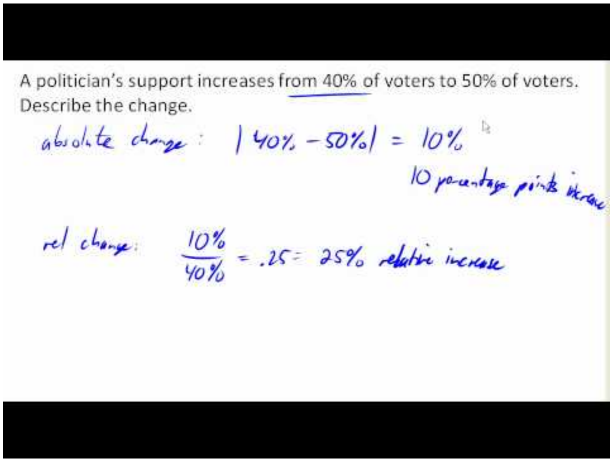
Example

A basketball player scores on 40% of 2-point field goal attempts, and on 30% of 3-point of field goal attempts. Find the player's overall field goal percentage.

Solution:

It is very tempting to average these values, and claim the overall average is 35%, but this is likely not correct, since most players make many more 2-point attempts than 3-point attempts. We don't actually have enough information to answer the question. Suppose the player attempted 200 2-point field goals and 100 3-point field goals. Then that player made 200(0.40) = 80 2-point shots and 100(0.30) = 30 3-point shots. Overall, they player made 110 shots out of 300, for a $\frac{110}{300} = 0.367 = 36.7\%$ overall field goal percentage.

For more information about these cautionary tales using percentages, view the following.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

Proportions and Rates

If you wanted to power the city of Lincoln, Nebraska using wind power, how many wind turbines would you need to install? Questions like these can be answered using rates and proportions.

Rates

A rate is the ratio (fraction) of two quantities.

A **unit rate** is a rate with a denominator of one.

Example

Your car can drive 300 miles on a tank of 15 gallons. Express this as a rate.

Solution:

Expressed as a rate, $\frac{300 \text{ miles}}{15 \text{ gallons}}$. We can divide to find a unit rate: $\frac{20 \text{ miles}}{1 \text{ gallon}}$, which we could also write as $20 \frac{\text{miles}}{\text{gallon}}$, or just 20 miles per gallon.[/hidden-answer]

Proportion Equation

A proportion equation is an equation showing the equivalence of two rates or ratios.

For an overview on rates and proportions, using the examples on this page, view the following video.

Click here to see the example.

Example

Solve the proportion $\frac{5}{3} = \frac{x}{6}$ for the unknown value *x*.

Solution:

This proportion is asking us to find a fraction with denominator 6 that is equivalent to the fraction $\frac{5}{3}$. We can solve this by multiplying both sides of the equation by 6, giving $x = \frac{5}{3} \cdot 6 = 10$.

Example

A map scale indicates that ½ inch on the map corresponds with 3 real miles. How many miles apart are two cities that are $2\dfrac{1}{4}$ inches apart on the map

Solution:

We can set up a proportion by setting equal two $\dfrac{\text{map inches}}{\text{real miles}}$ rates, and introducing a variable, *x*, to represent the unknown quantity—the mile distance between the cities.

$\dfrac{\frac{1}{2}\text{map inch}}{3\text{ miles}} = \dfrac{2\frac{1}{4}\text{map inches}}{x\text{ miles}}$	Multiply both sides by *x* and rewriting the mixed number
$\dfrac{\frac{1}{2}}{3} \cdot x = \dfrac{9}{4}$	Multiply both sides by 3
$\dfrac{1}{2}x = \dfrac{27}{4}$	Multiply both sides by 2 (or divide by ½)
$x = \dfrac{27}{2} = 13\dfrac{1}{2}\text{ miles}$	

Many proportion problems can also be solved using

dimensional analysis, the process of multiplying a quantity by rates to change the units.

Example

Your car can drive 300 miles on a tank of 15 gallons. How far can it drive on 40 gallons?

Solution:

We could certainly answer this question using a proportion: $\dfrac{300 \text{ miles}}{15 \text{ gallons}} = \dfrac{x \text{ miles}}{40 \text{ gallons}}$.

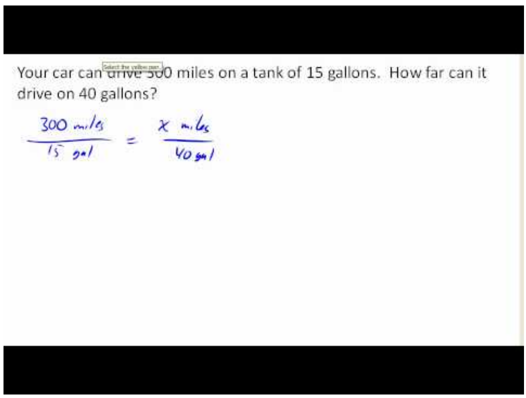
However, we earlier found that 300 miles on 15 gallons gives a rate of 20 miles per gallon. If we multiply the given 40 gallon quantity by this rate, the *gallons* unit "cancels" and we're left with a number of miles:

$$40 \text{ gallons} \cdot \frac{20 \text{ miles}}{\text{gallon}} = \frac{40 \text{ gallons}}{1} \cdot \frac{20 \text{ miles}}{\text{gallons}} = 800 \text{ miles}$$

Notice if instead we were asked "how many gallons are needed to drive 50 miles?" we could answer this question by inverting the 20 mile per gallon rate so that the *miles* unit cancels and we're left with gallons:

$$50\text{ miles}\cdot\frac{1\text{ gallon}}{20\text{ miles}}=\frac{50\text{ miles}}{1}\cdot\frac{1\text{ gallon}}{20\text{ miles}}=\frac{50\text{ gallons}}{20}=2.5\text{ gallons}$$

A worked example of this last question can be found in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

Notice that with the miles per gallon example, if we double the miles driven, we double the gas used. Likewise, with the map distance example, if the map distance doubles, the real-life distance doubles. This is a key feature of proportional relationships, and one we must confirm before assuming two things are related proportionally.

You have likely encountered distance, rate, and time problems in the past. This is likely because they are easy to visualize and most of us have experienced them first hand. In our next example, we will solve distance, rate and time problems that will require us to change the units that the distance or time is measured in.

Example

A bicycle is traveling at 15 miles per hour. How many feet will it cover in 20 seconds?

Solution:

To answer this question, we need to convert 20 seconds into feet. If we know the speed of the bicycle in feet per second, this question would be simpler. Since we don't, we will need to do additional unit conversions. We will need to know that 5280 ft = 1 mile. We might start by converting the 20 seconds into hours:

$$20 \text{ seconds} \cdot \frac{1 \text{ minute}}{60 \text{ seconds}} \cdot \frac{1 \text{ hour}}{60 \text{ minutes}} = \frac{1}{180} \text{ hour}$$

Now we can multiply by the 15 miles/hr

$$\frac{1}{180} \text{ hour} \cdot \frac{15 \text{ miles}}{1 \text{ hour}} = \frac{1}{12} \text{ mile}$$

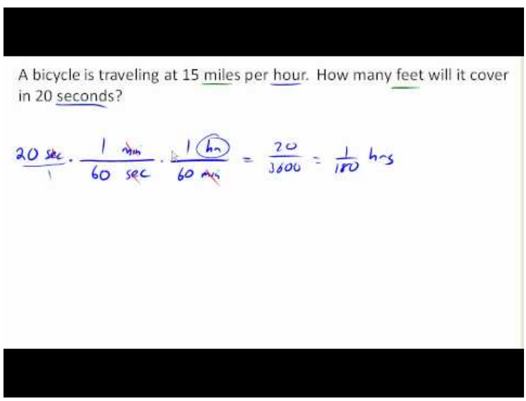
Now we can convert to feet

$$\frac{1}{12}\ \text{mile} \cdot \frac{5280\ \text{feet}}{1\ \text{mile}} = 440\ \text{feet}$$

We could have also done this entire calculation in one long set of products:

$$20\ \text{seconds} \cdot \frac{1\ \text{minute}}{60\ \text{seconds}} \cdot \frac{1\ \text{hour}}{60\ \text{minutes}} = \frac{15\ \text{miles}}{1\ \text{miles}} = \frac{5280\ \text{feet}}{1\ \text{mile}} = \frac{1}{180}\ \text{hour}$$

View the following video to see this problem worked through.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

TRY IT NOW

Click here to try!

Example

Suppose you're tiling the floor of a 10 ft by 10 ft room, and find that 100 tiles will be needed. How many tiles will be needed to tile the floor of a 20 ft by 20 ft room?

Solution:

In this case, while the width the room has doubled, the area has quadrupled. Since the number of tiles needed corresponds with the area of the floor, not the width, 400 tiles will be needed. We could find this using a proportion based on the areas of the rooms:

$$\frac{100 \text{ tiles}}{100\text{ft}^2} = \frac{n \text{ tiles}}{400\text{ft}^2}$$

Other quantities just don't scale proportionally at all.

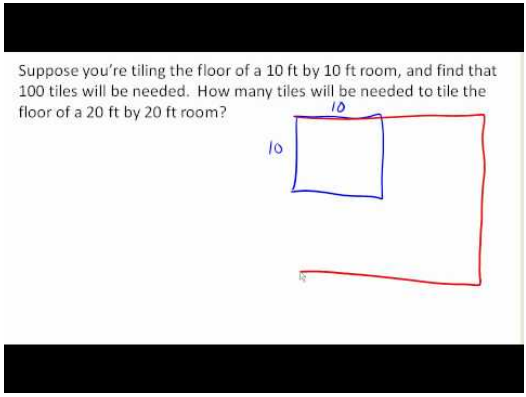
Example

Suppose a small company spends $1000 on an advertising campaign, and gains 100 new customers from it. How many new customers should they expect if they spend $10,000?

Solution:

While it is tempting to say that they will gain 1000 new customers, it is likely that additional advertising will be less effective than the initial advertising. For example, if the company is a hot tub store, there are likely only a fixed number of people interested in buying a hot tub, so there might not even be 1000 people in the town who would be potential customers.

Matters of scale in this example and the previous one are explained in more detail here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

Sometimes when working with rates, proportions, and percents, the process can be made more challenging by the magnitude of the numbers involved. Sometimes, large numbers are just difficult to comprehend.

Examples

The 2010 U.S. military budget was $683.7 billion. To gain perspective on how much money this is, answer the following questions.

1. What would the salary of each of the 1.4 million Walmart employees in the US be if the military budget were distributed evenly amongst them?
2. If you distributed the military budget of 2010 evenly amongst the 300 million people who live in the US, how much money would you give to each person?
3. If you converted the US budget into $100 bills, how long would it take you to count it out – assume it takes one second to count one $100 bill.

Solutions:

Here we have a very large number, about $683,700,000,000 written out. Of course, imagining a billion dollars is very difficult, so it can help to compare it to other quantities.

1. If that amount of money was used to pay the

salaries of the 1.4 million Walmart employees in the U.S., each would earn over $488,000.

2. There are about 300 million people in the U.S. The military budget is about $2,200 per person.
3. If you were to put $683.7 billion in $100 bills, and count out 1 per second, it would take 216 years to finish counting it.

Example

Compare the electricity consumption per capita in China to the rate in Japan.

Solution:

To address this question, we will first need data. From the CIA[3] website we can find the electricity consumption in 2011 for China was

3. https://www.cia.gov/library/publications/the-world-factbook/rankorder/2042rank.html

4,693,000,000,000 KWH (kilowatt-hours), or 4.693 trillion KWH, while the consumption for Japan was 859,700,000,000, or 859.7 billion KWH. To find the rate per capita (per person), we will also need the population of the two countries. From the World Bank,[4] we can find the population of China is 1,344,130,000, or 1.344 billion, and the population of Japan is 127,817,277, or 127.8 million.

Computing the consumption per capita for each country:

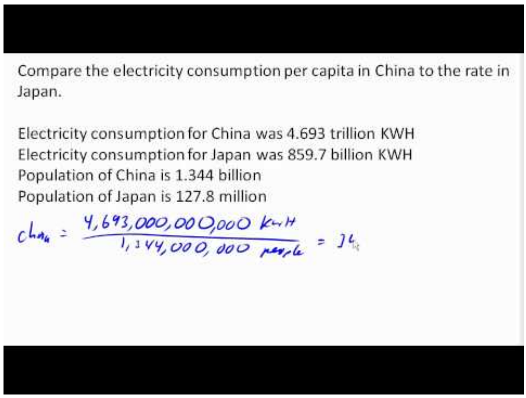
China: $\frac{4,693,000,000,000\text{KWH}}{1,344,130,000 \text{ people}}$ ≈ 3491.5 KWH per person

Japan: $\frac{859,700,000,000\text{KWH}}{127,817,277 \text{ people}}$ ≈ 6726 KWH per person

While China uses more than 5 times the electricity of Japan overall, because the population of Japan is so much smaller, it turns out Japan uses almost twice the electricity per person compared to China.

4. http://data.worldbank.org/indicator/SP.POP.TOTL

Working with large numbers is examined in more detail in this video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

A Bit of Geometry

Geometric shapes, as well as area and volumes, can often be important in problem solving.

Let's start things off with an example, rather than trying to explain geometric concepts to you.

Example

You are curious how tall a tree is, but don't have any way to climb it. Describe a method for determining the height.

Solution:

There are several approaches we could take. We will use one based on triangles, which requires that it's a sunny day. Suppose the tree is casting a shadow, say 15 ft long. I can then have a friend help me measure my own shadow. Suppose I am 6 ft tall, and case a 1.5 ft shadow. Since the triangle formed by the tree and its shadow has the same angles as the triangle formed by me and my shadow, these triangles are called ***similar triangles*** and their sides will scale proportionally. In other words, the ratio of height to width will be the same in both triangles. Using this, we can find the height of the tree, which we'll denote by ***h***:

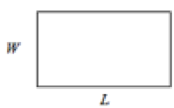
$$\frac{6}{1.5} = \frac{h}{15}$$

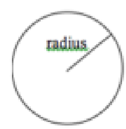
If we multiply each side by 15, h, or the height of the tree, is approximately 60 feet tall.

It may be helpful to recall some formulas for areas and volumes of a few basic shapes:

Areas

Rectangle	**Circ**
Area: $L \times W$	Area:
Perimeter: 2L +2W	Circur





Rectangular Box	Cylinder
Volume: $L \times W \times H$	Volume: $\pi r^2 h$

In our next two examples, we will combine the ideas we have explored about ratios with the geometry of some basic shapes to answer questions. In the first example, we will predict how much dough will be needed for a pizza that is 16 inches in diameter given that we know how much dough it takes for a pizza with a diameter of 12 inches. The second example uses the volume of a cylinder to determine the number of calories in a marshmallow.

Examples

If a 12 inch diameter pizza requires 10 ounces of dough, how much dough is needed for a 16 inch pizza?

Solution:

To answer this question, we need to consider how the weight of the dough will scale. The weight will be based on the volume of the dough. However, since both pizzas will be about the same thickness, the weight will scale with the area of the top of the pizza. We can find the area of each pizza using the formula for area of a circle, $A = \pi r^2$:

A 12" pizza has radius 6 inches, so the area will be $\pi 6^2$ = about 113 square inches.

A 16" pizza has radius 8 inches, so the area will be $\pi 8^2$ = about 201 square inches.

Notice that if both pizzas were 1 inch thick, the volumes would be 113 in^3 and 201 in^3 respectively, which are at the same ratio as the areas. As mentioned earlier, since the thickness is the same for both pizzas, we can safely ignore it.

We can now set up a proportion to find the weight of the dough for a 16" pizza:

$$\frac{10 \text{ ounces}}{113\text{in}^2} = \frac{x \text{ ounces}}{201\text{in}^2}$$

Multiply both sides by 201

$x = 201 \cdot \frac{10}{113}$ = about 17.8 ounces of dough for a 16" pizza.

It is interesting to note that while the diameter is $\frac{16}{12}$ = 1.33 times larger, the dough required, which scales with area, is 1.33^2 = 1.78 times larger.

The following video illustrates how to solve this problem.

[Click here to view this video.]

Example

A company makes regular and jumbo marshmallows. The regular marshmallow has 25 calories. How many calories will the jumbo marshmallow have?

Solution:

We would expect the calories to scale with volume. Since the marshmallows have cylindrical shapes, we can use that formula to find the volume. From the grid in the image, we can estimate the radius and height of each marshmallow.

The regular marshmallow appears to have a diameter of about 3.5 units, giving a radius of 1.75 units, and a height of about 3.5 units. The volume is about $\pi(1.75)^2(3.5) = 33.7$ units^3.

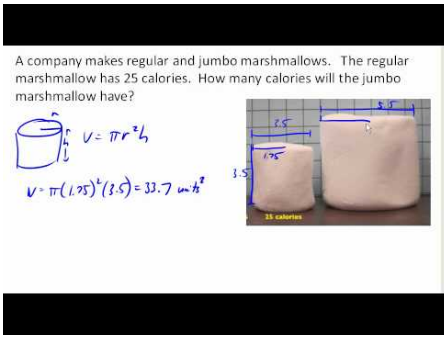
The jumbo marshmallow appears to have a diameter of about 5.5 units, giving a radius of 2.75 units, and a height of about 5 units. The volume is about $\pi(2.75)^2(5) = 118.8$ units^3.

We could now set up a proportion, or use rates. The regular marshmallow has 25 calories for 33.7 cubic units of volume. The jumbo marshmallow will have:

$$118.8 \text{ units}^3 \cdot \frac{25 \text{ calories}}{33.7 \text{ units}^3} = 88.1 \text{ calories}$$

It is interesting to note that while the diameter and height are about 1.5 times larger for the jumbo marshmallow, the volume and calories are about $1.5^3 = 3.375$ times larger.

For more about the marshmallow example, watch this video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=61

TRY IT NOW

A website says that you'll need 48 fifty-pound bags of sand to fill a sandbox that measure 8ft by

8ft by 1ft. How many bags would you need for a sandbox 6ft by 4ft by 1ft?

Solution:

The original sandbox has volume 64ft^3. The smaller sandbox has volume 24ft^3.

$\dfrac{48\text{bags}}{64\text{ft}^2} = \dfrac{x\ \text{bags}}{24\text{in}^3}$ results in x = 18 bags.

Mary (from the application that started this topic), decides to use what she knows about the height of the roof to measure the height of her second daughter. If her second daughter casts a shadow that is 1.5 feet long when she is 13.5 feet from the house, what is the height of the second daughter? Draw an accurate diagram and use similar triangles to solve.

Solution:

2.5 ft

Click here to view the solution in a video.

In the next section, we will explore the process of combining different types of information to answer questions.

Attributions

SOLVING PROBLEMS WITH MATH

Learning Outcomes

- Identify and apply a solution pathway for multi-step problems

In this section we will bring together the mathematical tools we've reviewed, and use them to approach more complex problems. In many problems, it is tempting to take the given information, plug it into whatever formulas you have handy, and hope that the result is what you were supposed to find. Chances are, this approach has served you well in other math classes.

This approach does not work well with real life problems, however. Read on to learn how to use a generalized problem solving approach to solve a wide variety of quantitative problems, including how taxes are calculated.

Problem Solving and Estimating

Problem solving is best approached by first starting at the end: identifying exactly what you are looking for. From there, you then work backwards, asking "what information and procedures will I need to find this?" Very few interesting questions can be answered in one mathematical step; often times you will need to chain together a **solution pathway**, a series of steps that will allow you to answer the question.

Problem Solving Process

1. Identify the question you're trying to answer.
2. Work backwards, identifying the information you will need and the relationships you will use to answer that question.
3. Continue working backwards, creating a solution pathway.
4. If you are missing necessary information, look it up or estimate it. If you have unnecessary information, ignore it.
5. Solve the problem, following your solution

pathway.

In most problems we work, we will be approximating a solution, because we will not have perfect information. We will begin with a few examples where we will be able to approximate the solution using basic knowledge from our lives.

In the first example, we will need to think about time scales, we are asked to find how many times a heart beats in a year, but usually we measure heart rate in beats per minute.

Examples

How many times does your heart beat in a year?

Solution:

This question is asking for the rate of heart beats per year. Since a year is a long time to measure heart beats for, if we knew the rate of heart beats per minute, we could scale that quantity up to a year. So the information we need to answer this question is heart beats per minute. This is something you can easily measure by counting your pulse while watching a clock for a minute.

Suppose you count 80 beats in a minute. To convert this to beats per year:

$$\frac{80 \text{ beats}}{1 \text{ minute}} \cdot \frac{60 \text{ minutes}}{1 \text{ hour}} \cdot \frac{24 \text{ hours}}{1 \text{ day}} \cdot \frac{365 \text{ days}}{1 \text{ year}} = 42,048,000 \text{ beats per year}$$

The technique that helped us solve the last problem was to get the number of heartbeats in a minute translated into the number of heartbeats in a year. Converting units from one to another, like minutes to years is a common tool for solving problems.

In the next example, we show how to infer the thickness of something too small to measure with every-day tools. Before precision instruments were widely available, scientists and engineers had to get creative with ways to measure either very small or very large things. Imagine how early astronomers inferred the distance to stars, or the circumference of the earth.

Example

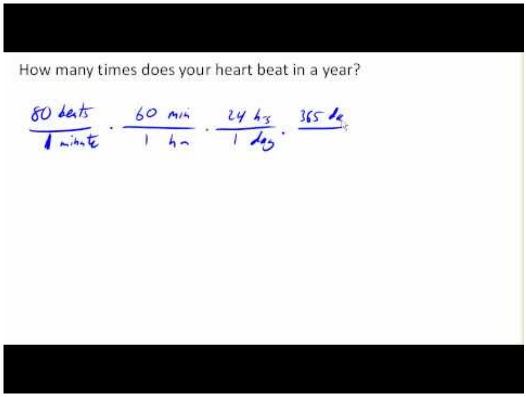
How thick is a single sheet of paper? How much does it weigh?

Solution:
While you might have a sheet of paper handy, trying to measure it would be tricky. Instead we might imagine a stack of paper, and then scale the thickness and weight to a single sheet. If you've ever bought paper for a printer or copier, you probably bought a ream, which contains 500 sheets. We could estimate that a ream of paper is about 2 inches thick and weighs about 5 pounds. Scaling these down,

$$\frac{2 \text{ inches}}{\text{ream}} \cdot \frac{1 \text{ ream}}{500 \text{ pages}} = 0.004 \text{ inches per sheet}$$

$$\frac{5 \text{ pounds}}{\text{ream}} \cdot \frac{1 \text{ ream}}{500 \text{ pages}} = 0.01 \text{ pounds per sheet, or } = 0.16 \text{ ounces per sheet.}$$

The first two example questions in this set are examined in more detail here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=63

We can infer a measurement by using scaling. If 500 sheets of paper is two inches thick, then we could use proportional reasoning to infer the thickness of one sheet of paper.

In the next example, we use proportional reasoning to determine how many calories are in a mini muffin when you are given the amount of calories for a regular sized muffin.

Example

A recipe for zucchini muffins states that it yields 12 muffins, with 250 calories per muffin. You instead decide to make mini-muffins, and the recipe yields 20 muffins. If you eat 4, how many calories will you consume?

Solution:

There are several possible solution pathways to answer this question. We will explore one.

To answer the question of how many calories 4 mini-muffins will contain, we would want to know the number of calories in each mini-muffin. To find the calories in each mini-muffin, we could first find the total calories for the entire recipe, then divide it by the number of mini-muffins produced. To find the total calories for the recipe, we could multiply the calories per standard muffin by the number per muffin. Notice that this produces a multi-step solution pathway. It is often easier to solve a problem in small steps, rather than trying to find a way to jump directly from the given information to the solution.

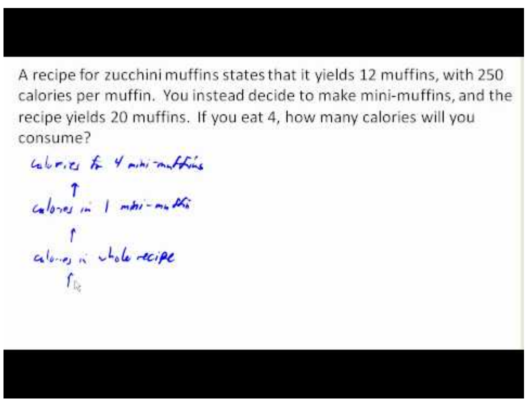
We can now execute our plan:

$$12 \text{ muffins} \cdot \frac{250 \text{ calories}}{\text{muffin}} = 3000 \text{ calories for the whole recipe}$$

$$\frac{3000 \text{ calories}}{20 \text{ mini-muffins}} = \text{ gives } 150 \text{ calories per mini-muffin}$$

$$4 \text{ mini-muffins} \cdot \frac{150 \text{ calories}}{\text{mini-muffin}} = \text{totals } 600 \text{ calories consumed.}$$

View the following video for more about the zucchini muffin problem.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=63

We have found that ratios are very helpful when we know some information but it is not in the right units, or parts to answer our question. Making comparisons mathematically often involves using ratios and proportions. For the last

Example

You need to replace the boards on your deck. About how much will the materials cost?

Solution:

There are two approaches we could take to this problem: 1) estimate the number of boards we will need and find the cost per board, or 2) estimate the area of the deck and find the approximate cost per square foot for deck boards. We will take the latter approach.

For this solution pathway, we will be able to answer the question if we know the cost per square foot for decking boards and the square footage of the deck. To find the cost per square foot for decking boards, we could compute the area of a single board, and divide it into the cost for that board. We can compute the square footage of the deck using geometric formulas. So first we need information: the dimensions of the deck, and the cost and dimensions of a single deck board.

Suppose that measuring the deck, it is rectangular, measuring 16 ft by 24 ft, for a total area of 384 ft^2.

From a visit to the local home store, you find that an 8 foot by 4 inch cedar deck board costs about \$7.50. The area of this board, doing the necessary conversion from inches to feet, is:

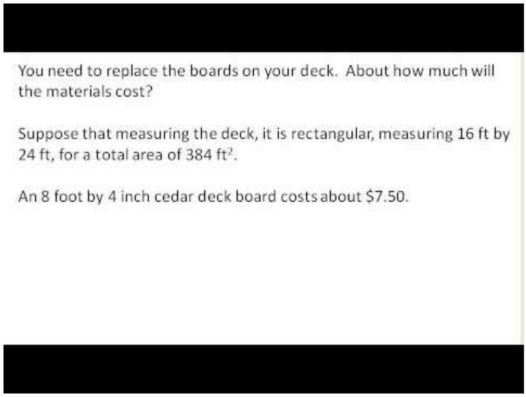
$8 \text{ feet} \cdot 4 \text{ inches} \cdot \dfrac{1 \text{ foot}}{12 \text{ inches}} = 2.667\text{ft}^2$. The cost per square foot is then $\dfrac{\$7.50}{2.667\text{ft}^2} = \$2.8125 \text{ per ft}^2$.

This will allow us to estimate the material cost for the whole 384 ft^2 deck

$\$384\text{ft}^2 \cdot \dfrac{\$2.8125}{\text{ft}^2} = \$1080 \text{ total cost}.$

Of course, this cost estimate assumes that there is no waste, which is rarely the case. It is common to add at least 10% to the cost estimate to account for waste.

This example is worked through in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=63

Example

Is it worth buying a Hyundai Sonata hybrid instead the regular Hyundai Sonata?

Solution:

To make this decision, we must first decide what our basis for comparison will be. For the purposes of this example, we'll focus on fuel and purchase costs, but environmental impacts and maintenance costs are other factors a buyer might consider.

It might be interesting to compare the cost of gas to run both cars for a year. To determine this, we will need to know the miles per gallon both cars get, as well as the number of miles we expect to drive in a year. From that information, we can find the number of gallons required from a year. Using the price of gas per gallon, we can find the running cost.

From Hyundai's website, the 2013 Sonata will get 24 miles per gallon (mpg) in the city, and 35 mpg on the highway. The hybrid will get 35 mpg in the city, and 40 mpg on the highway.

An average driver drives about 12,000 miles a year. Suppose that you expect to drive about 75% of that in the city, so 9,000 city miles a year, and 3,000 highway miles a year.

We can then find the number of gallons each car would require for the year.

Sonata:

$$9000\text{ city miles}\cdot\frac{1\text{ gallon}}{24\text{ city miles}}+3000\text{ highway miles}\cdot\frac{1\text{ gallon}}{35\text{ highway miles}}=460.7\text{ gallons}$$

Hybrid:

$9000 \text{ city miles} \cdot \frac{1 \text{ gallon}}{35 \text{ city miles}} + 3000 \text{ highway miles} \cdot \frac{1 \text{ gallon}}{40 \text{ highway miles}} = 332.1 \text{ gallons}$

If gas in your area averages about $3.50 per gallon, we can use that to find the running cost:

Sonata: $460.7 \text{ gallons} \cdot \dfrac{\$3.50}{\text{gallon}} = \$1612.45$

Hybrid: $332.1 \text{ gallons} \cdot \dfrac{\$3.50}{\text{gallon}} = \$1162.35$

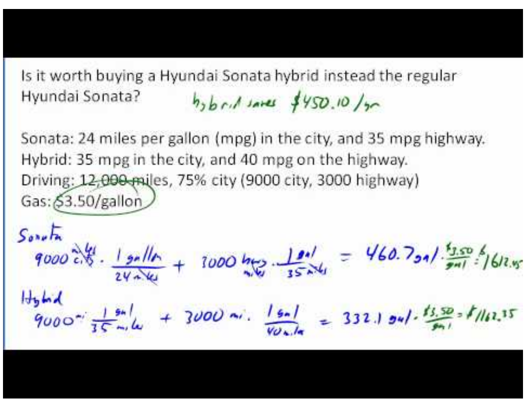
The hybrid will save $450.10 a year. The gas costs for the hybrid are about $\dfrac{\$450.10}{\$1612.45}$ = 0.279 = 27.9% lower than the costs for the standard Sonata.

While both the absolute and relative comparisons are useful here, they still make it hard to answer the original question, since “is it worth it” implies there is some tradeoff for the gas savings. Indeed, the hybrid Sonata costs about $25,850, compared to the base model for the regular Sonata, at $20,895.

To better answer the “is it worth it” question, we might explore how long it will take the gas savings to make up for the additional initial cost. The hybrid costs $4965 more. With gas savings of $451.10 a year, it will take about 11 years for the gas savings to make up for the higher initial costs.

We can conclude that if you expect to own the car 11 years, the hybrid is indeed worth it. If you plan to own the car for less than 11 years, it may still be worth it, since the resale value of the hybrid may be higher, or for other non-monetary reasons. This is a case where math can help guide your decision, but it can't make it for you.

This question pulls together all the skills discussed previously on this page, as the video demonstration illustrates.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=63

Try It

#1. Click here to try this problem.

#2. Click here to try this problem.

#3. Click here to try this problem.

Taxes

Governments collect taxes to pay for the services they provide.

In the United States, federal income taxes help fund the military, the environmental protection agency, and thousands of other programs. Property taxes help fund schools. Gasoline taxes help pay for road improvements. While very few people enjoy paying taxes, they are necessary to pay for the services we all depend upon.

Taxes can be computed in a variety of ways, but are typically computed as a percentage of a sale, of one's income, or of one's assets.

Example: Sales Tax

The sales tax rate in a city is 9.3%. How much sales tax will you pay on a \$140 purchase?

Solution:

The sales tax will be 9.3% of \$140. To compute this, we multiply \$140 by the percent written as a decimal: \$140(0.093) = \$13.02.

Click here to try another version of this problem:

When taxes are not given as a fixed percentage rate, sometimes it is necessary to calculate the **effective tax**

rate: the equivalent percent rate of the tax paid out of the dollar amount the tax is based on.

Example: Property Tax

Jaquim paid $3,200 in property taxes on his house valued at $215,000 last year. What is the effective tax rate?

Solution:

We can compute the equivalent percentage: 3200/215000 = 0.01488, or about 1.49% effective rate.

Taxes are often referred to as progressive, regressive, or flat.

- A **flat tax**, or proportional tax, charges a constant percentage rate.
- A **progressive tax** increases the percent rate as the base amount increases.
- A **regressive tax** decreases the percent rate as the base amount increases.

Example: Federal Income Tax

The United States federal income tax on earned wages is an example of a progressive tax. People with a higher wage income pay a higher percent tax on their income.

For a single person in 2011, adjusted gross income (income after deductions) under $8,500 was taxed at 10%. Income over $8,500 but under $34,500 was taxed at 15%.

Earning $10,000

Stephen earned $10,000 in 2011. He would pay 10% on the portion of his income under $8,500, and 15% on the income over $8,500.

8500(0.10) = 850   10% of $8500
1500(0.15) = 225   15% of the remaining $1500 of income
Total tax: = $1075

What was Stephen's effective tax rate?

Solution:

The effective tax rate paid is 1075/10000 = 10.75%

Example: Gasoline Tax

A gasoline tax is a flat tax when considered in terms of consumption. A tax of, say, $0.30 per gallon is proportional to the amount of gasoline purchased. Someone buying 10 gallons of gas at $4 a gallon would pay $3 in tax, which is $3/$40 = 7.5%. Someone buying 30 gallons of gas at $4 a gallon would pay $9 in tax, which is $9/$120 = 7.5%, the same effective rate.

However, in terms of income, a gasoline tax is often considered a regressive tax. It is likely that someone earning $30,000 a year and someone earning $60,000 a year will drive about the same amount. If both pay $60 in gasoline taxes over a year, the person earning $30,000 has paid 0.2% of their income, while the person earning $60,000 has paid 0.1% of their income in gas taxes.

Try It

A sales tax is a fixed percentage tax on a person's purchases. Is this a flat, progressive, or regressive tax?

Solution:

While sales tax is a flat percentage rate, it is often considered a regressive tax for the same reasons as the gasoline tax.

Click here to try other tax problems.

Attributions

PUTTING IT TOGETHER: GENERAL PROBLEM SOLVING

Now that you have seen a number of general problem solving techniques and plenty of examples in this module, let's try to put it all together. In real life, sometimes we have to make big decisions. A major career change or relocation can throw your life into disarray, but perhaps even more importantly, such a change can have lasting effects on your financial future.

Suppose you are currently working as a bank teller in your hometown. Your job pays \$11 per hour, with 10% taxes taken out of each paycheck. For simplicity, assume you have to work 260 days of the year (you don't work weekends), 8 hours a day. By carefully budgeting you can keep you monthly expenses to about \$1500.

But now another job offer has come up. You have the opportunity to become an assistant manager. The only catch is that you'll have to move to a new branch of the bank opening up in a nearby metropolitan area. So here are the details:

The assistant manager salary starts at \$32,000. This puts you into the next tax bracket, which means that you will pay \$2500 plus 15% of the amount earned over \$25,000. Because you are moving to a bigger city, your living expenses will rise as well. After a little research, you determine that your monthly expenses will probably be around \$2000.

Should you make the move and take the job offer?

This is a tough decision! We should do some calculations first. How much money do you make in a full year as a bank teller? First let's find out how many total hours you work in a year.

$$8 \text{ hrs./day} \times 260 \text{ days} = 2080 \text{ hrs.}$$

Next, multiply by the hourly wage to determine your **gross income**.

$$2080 \text{ hrs.} \times 11 \text{ \$/hr.} = \$22{,}880$$

But don't forget about taxes! You will have to pay 10% (or 0.10) of this amount to the US government.

$22,880 x 0.10=$2288

After subtracting the tax, this leaves you with your **net income**, or **take-home pay**:

$22,880-$2288=$20,592

Ok, so now let's figure out the yearly expenses. If monthly expenses are $1500, then each year you will pay:

12 x $1500=$18,000

Finally, after the bills are paid, you can do what we want to with the remainder. This is our **discretionary budget**, and it is as good a measure as any as to how successful you are.

$20,592 – $18,000 = $2592

Now keep that number $2592 in mind. Let's see what kind of discretionary budget you will have if you take the new job. Since salary is, by definition, a yearly income amount, our first task is to compute and deduct the taxes. This time, you will pay $2500 plus 15% of the difference between the salary and $25,000.

$2500+($7000 x 0.15)=$3550

Therefore, the net income would be:

$32,000-$3550=$28,450

It's a higher net income than you are currently making, but how will it stack up against the higher cost-of-living expenses in the city? Take the estimated monthly expenses of $2000 and multiply it by 12 months:

12 x $2000=$24,000

Thus your new discretionary budget would be:

$28,450-$24,000=$4450

That's definitely an improvement over $2592. Maybe it's time to move to the big city and start advancing your career. However, you can see that it's not really that much more, so you probably shouldn't go out and buy a brand new car. Just wait until you get your first promotion to full manager.

Attributions

PROBLEM SOLVING EXERCISES

Exercises

1. Out of 230 racers who started the marathon, 212 completed the race, 14 gave up, and 4 were disqualified. What percentage did not complete the marathon?

2. Patrick left an $8 tip on a $50 restaurant bill. What percent tip is that?

3. Ireland has a 23% VAT (value-added tax, similar to a sales tax). How much will the VAT be on a purchase of a €250 item?

4. Employees in 2012 paid 4.2% of their gross wages towards social security (FICA tax), while employers paid

another 6.2%. How much will someone earning $45,000 a year pay towards social security out of their gross wages?

5. A project on Kickstarter.com was aiming to raise $15,000 for a precision coffee press. They ended up with 714 supporters, raising 557% of their goal. How much did they raise?

6. Another project on Kickstarter for an iPad stylus raised 1,253% of their goal, raising a total of $313,490 from 7,511 supporters. What was their original goal?

7. The population of a town increased from 3,250 in 2008 to 4,300 in 2010. Find the absolute and relative (percent) increase.

8. The number of CDs sold in 2010 was 114 million, down from 147 million the previous year[6]. Find the absolute and relative (percent) decrease.

9. A company wants to decrease their energy use by 15%.
    1. If their electric bill is currently $2,200 a month, what will their bill be if they're successful?
    2. If their next bill is $1,700 a month, were they successful? Why or why not?

10. A store is hoping an advertising campaign will increase their number of customers by 30%. They currently have about 80 customers a day.
    1. How many customers will they have if their campaign is successful?
    2. If they increase to 120 customers a day, were they successful? Why or why not?

11. An article reports "attendance dropped 6% this year, to 300." What was the attendance before the drop?

12. An article reports "sales have grown by 30% this year, to $200 million." What were sales before the growth?

13. The Walden University had 47,456 students in 2010, while Kaplan University had 77,966 students. Complete the following statements:
    1. Kaplan's enrollment was ___% larger than Walden's.
    2. Walden's enrollment was ___% smaller than Kaplan's.
    3. Walden's enrollment was ___% of Kaplan's.

14. In the 2012 Olympics, Usain Bolt ran the 100m dash in 9.63 seconds. Jim Hines won the 1968 Olympic gold with a time of 9.95 seconds.
    1. Bolt's time was ___% faster than Hines'.
    2. Hine' time was ___% slower than Bolt's.
    3. Hine' time was ___% of Bolt's.

15. A store has clearance items that have been marked down by 60%. They are having a sale, advertising an additional 30% off clearance items. What percent of the original price do you end up paying?

16. Which is better: having a stock that goes up 30% on Monday than drops 30% on Tuesday, or a stock that drops 30% on Monday and goes up 30% on Tuesday? In each case, what is the net percent gain or loss?

17. Are these two claims equivalent, in conflict, or not comparable because they're talking about different things?
    16. "16.3% of Americans are without health insurance"[7]
    17. "only 55.9% of adults receive employer provided health insurance"[8]

18. Are these two claims equivalent, in conflict, or not comparable because they're talking about different things?
    1. "We mark up the wholesale price by 33% to come up with the retail price"
    2. "The store has a 25% profit margin"

19. Are these two claims equivalent, in conflict, or not comparable because they're talking about different things?
    1. "Every year since 1950, the number of American children gunned down has doubled."
    2. "The number of child gunshot deaths has doubled from 1950 to 1994."

20. Are these two claims equivalent, in conflict, or not comparable because they're talking about different things?[9]
    1. "75 percent of the federal health care law's taxes would be paid by those earning less than $120,000 a year"
    2. "76 percent of those who would pay the penalty [health care law's taxes] for not having insurance in 2016 would earn under $120,000"

21. Are these two claims equivalent, in conflict, or not comparable because they're talking about different things?
    1. "The school levy is only a 0.1% increase of the property tax rate."

2. “This new levy is a 12% tax hike, raising our total rate to $9.33 per $1000 of value.”

22. Are the values compared in this statement comparable or not comparable? “Guns have murdered more Americans here at home in recent years than have died on the battlefields of Iraq and Afghanistan. In support of the two wars, more than 6,500 American soldiers have lost their lives. During the same period, however, guns have been used to murder about 100,000 people on American soil”[10]

23. A high school currently has a 30% dropout rate. They’ve been tasked to decrease that rate by 20%. Find the equivalent percentage point drop.

24. A politician’s support grew from 42% by 3 percentage points to 45%. What percent (relative) change is this?

25. Marcy has a 70% average in her class going into the final

exam. She says "I need to get a 100% on this final so I can raise my score to 85%." Is she correct?

26. Suppose you have one quart of water/juice mix that is 50% juice, and you add 2 quarts of juice. What percent juice is the final mix?

27. Find a unit rate: You bought 10 pounds of potatoes for $4.

28. Find a unit rate: Joel ran 1500 meters in 4 minutes, 45 seconds.

29. Solve: $\frac{2}{5} = \frac{6}{x}$.

30. Solve: $\frac{n}{5} = \frac{16}{20}$.

31. A crepe recipe calls for 2 eggs, 1 cup of flour, and 1 cup of milk. How much flour would you need if you use 5 eggs?

32. An 8ft length of 4 inch wide crown molding costs $14. How much will it cost to buy 40ft of crown molding?

33. Four 3-megawatt wind turbines can supply enough electricity to power 3000 homes. How many turbines would be required to power 55,000 homes?

34. A highway had a landslide, where 3,000 cubic yards of material fell on the road, requiring 200 dump truck loads to clear. On another highway, a slide left 40,000 cubic yards on the road. How many dump truck loads would be needed to clear this slide?

35. Convert 8 feet to inches.

36. Convert 6 kilograms to grams.

37. A wire costs $2 per meter. How much will 3 kilometers of wire cost?

38. Sugar contains 15 calories per teaspoon. How many calories are in 1 cup of sugar?

39. A car is driving at 100 kilometers per hour. How far does it travel in 2 seconds?

40. A chain weighs 10 pounds per foot. How many ounces will 4 inches weigh?

41. The table below gives data on three movies. Gross earnings is the amount of money the movie brings in.

Compare the net earnings (money made after expenses) for the three movies.[11]

Movie	Release Date	Budget	Gross earnings
Saw	10/29/2004	$1,200,000	$103,096,345
Titanic	12/19/1997	$200,000,000	$1,842,879,955
Jurassic Park	6/11/1993	$63,000,000	$923,863,984

42. For the movies in the previous problem, which provided the best return on investment?

43. The population of the U.S. is about 309,975,000, covering a land area of 3,717,000 square miles. The population of India is about 1,184,639,000, covering a land area of 1,269,000 square miles. Compare the population densities of the two countries.

44. The GDP (Gross Domestic Product) of China was

$5,739 billion in 2010, and the GDP of Sweden was $435 billion. The population of China is about 1,347 million, while the population of Sweden is about 9.5 million. Compare the GDP per capita of the two countries.

45. In June 2012, Twitter was reporting 400 million tweets per day. Each tweet can consist of up to 140 characters (letter, numbers, etc.). Create a comparison to help understand the amount of tweets in a year by imagining each character was a drop of water and comparing to filling something up.

46. The photo sharing site Flickr had 2.7 billion photos in June 2012. Create a comparison to understand this number by assuming each picture is about 2 megabytes in size, and comparing to the data stored on other media like DVDs, iPods, or flash drives.

47. Your chocolate milk mix says to use 4 scoops of mix for 2 cups of milk. After pouring in the milk, you start adding

the mix, but get distracted and accidentally put in 5 scoops of mix. How can you adjust the mix if:

1. There is still room in the cup?
2. The cup is already full?

48. A recipe for sabayon calls for 2 egg yolks, 3 tablespoons of sugar, and ¼ cup of white wine. After cracking the eggs, you start measuring the sugar, but accidentally put in 4 tablespoons of sugar. How can you compensate?

49. The Deepwater Horizon oil spill resulted in 4.9 million barrels of oil spilling into the Gulf of Mexico. Each barrel of oil can be processed into about 19 gallons of gasoline. How many cars could this have fueled for a year? Assume an average car gets 20 miles to the gallon, and drives about 12,000 miles in a year.

50. The store is selling lemons at 2 for $1. Each yields about 2 tablespoons of juice. How much will it cost to buy enough lemons to make a 9-inch lemon pie requiring ½ cup of lemon juice?

51. A piece of paper can be made into a cylinder in two ways: by joining the short sides together, or by joining the long sides together[12]. Which cylinder would hold more? How much more?

52. Which of these glasses contains more liquid? How much more?

In the next 4 questions, estimate the values by making reasonable approximations for unknown values, or by doing some research to find reasonable values.

53. Estimate how many gallons of water you drink in a year.

54. Estimate how many times you blink in a day.

55. How much does the water in a 6-person hot tub weigh?

56. How many gallons of paint would be needed to paint a two-story house 40 ft long and 30 ft wide?

57. During the landing of the Mars Science Laboratory *Curiosity*, it was reported that the signal from the rover would take 14 minutes to reach earth. Radio signals travel at the speed of light, about 186,000 miles per second. How far was Mars from Earth when *Curiosity* landed?
58. It is estimated that a driver takes, on average, 1.5 seconds from seeing an obstacle to reacting by applying the brake or swerving. How far will a car traveling at 60 miles per hour travel (in feet) before the driver reacts to an obstacle?

59. The flash of lightning travels at the speed of light, which is about 186,000 miles per second. The sound of lightning (thunder) travels at the speed of sound, which

is about 750 miles per hour.

1. If you see a flash of lightning, then hear the thunder 4 seconds later, how far away is the lightning?
2. Now let's generalize that result. Suppose it takes *n* seconds to hear the thunder after a flash of lightning. How far away is the lightning, in terms of *n*?

60. Sound travels about 750 miles per hour. If you stand in a parking lot near a building and sound a horn, you will hear an echo.
    1. Suppose it takes about ½ a second to hear the echo. How far away is the building[13]?
    2. Now let's generalize that result. Suppose it takes *n* seconds to hear the echo. How far away is the building, in terms of *n*?

61. It takes an air pump 5 minutes to fill a twin sized air mattress (39 by 8.75 by 75 inches). How long will it take to fill a queen sized mattress (60 by 8.75 by 80 inches)?

62. It takes your garden hose 20 seconds to fill your 2-gallon watering can. How long will it take to fill
63. An inflatable pool measuring 3 feet wide, 8 feet long, and 1 foot deep.[14]
64. A circular inflatable pool 13 feet in diameter and 3 feet deep.[15]

63. You want to put a 2″ thick layer of topsoil for a new 20'x30′ garden. The dirt store sells by the cubic yards. How many cubic yards will you need to order?

64. A box of Jell-O costs $0.50, and makes 2 cups. How much would it cost to fill a swimming pool 4 feet deep, 8 feet wide, and 12 feet long with Jell-O? (1 cubic foot is about 7.5 gallons)

65. You read online that a 15 ft by 20 ft brick patio would cost about $2,275 to have professionally installed. Estimate the cost of having a 18 by 22 ft brick patio installed.
66. I was at the store, and saw two sizes of avocados being

sold. The regular size sold for $0.88 each, while the jumbo ones sold for $1.68 each. Which is the better deal?

67. The grocery store has bulk pecans on sale, which is great since you're planning on making 10 pecan pies for a wedding. Your recipe calls for 1¾ cups pecans per pie. However, in the bulk section there's only a scale available, not a measuring cup. You run over to the baking aisle and find a bag of pecans, and look at the nutrition label to gather some info. How many pounds of pecans should you buy?

68. Soda is often sold in 20 ounce bottles. The nutrition label for one of these bottles is shown to the right. A packet of sugar (the kind they have at restaurants for your coffee or tea) typically contain 4 grams of sugar in the U.S. Drinking a 20 oz soda is equivalent to eating how many packets of sugar?[16]

For the next set of questions, *first* identify the information you need to answer the question, and *then* turn to the end of the section to find that information. The details may be imprecise; answer the question the best you can with the provided information. Be sure to justify your decision.

69. You're planning on making 6 meatloafs for a party. You go to the store to buy breadcrumbs, and see they are sold by the canister. How many canisters do you need to buy?

70. Your friend wants to cover their car in bottle caps, like in this picture.[17] How many bottle caps are you going to need?

71. You need to buy some chicken for dinner tonight. You found an ad showing that the store across town has it on sale for $2.99 a pound, which is cheaper than your usual

neighborhood store, which sells it for $3.79 a pound. Is it worth the extra drive?

72. I have an old gas furnace, and am considering replacing it with a new, high efficiency model. Is upgrading worth it?

73. Janine is considering buying a water filter and a reusable water bottle rather than buying bottled water. Will doing so save her money?

74. Marcus is considering going car-free to save money and be more environmentally friendly. Is this financially a good decision?

For the next set of problems, research or make educated estimates for any unknown quantities needed to answer the question.

75. You want to travel from Tacoma, WA to Chico, CA for a wedding. Compare the costs and time involved with driving, flying, and taking a train. Assume that if you fly or take the train you'll need to rent a car while you're there. Which option is best?

76. You want to paint the walls of a 6ft by 9ft storage room that has one door and one window. You want to put on two coats of paint. How many gallons and/or quarts of paint should you buy to paint the room as cheaply as possible?

77. A restaurant in New York tiled their floor with pennies[18]. Just for the materials, is this more expensive than using a more traditional material like ceramic tiles? If each penny has to be laid by hand, estimate how long

it would take to lay the pennies for a 12ft by 10ft room. Considering material and labor costs, are pennies a cost-effective replacement for ceramic tiles?

78. You are considering taking up part of your back yard and turning it into a vegetable garden, to grow broccoli, tomatoes, and zucchini. Will doing so save you money, or cost you more than buying vegetables from the store?

79. Barry is trying to decide whether to keep his 1993 Honda Civic with 140,000 miles, or trade it in for a used 2008 Honda Civic. Consider gas, maintenance, and insurance costs in helping him make a decision.

80. Some people claim it costs more to eat vegetarian, while some claim it costs less. Examine your own grocery habits, and compare your current costs to the costs of switching your diet (from omnivore to vegetarian or vice versa as appropriate). Which diet is more cost effective based on your eating habits?

Info for the breadcrumbs question

Nutrition Facts

Serving Size: 1/3 cup (30g)
Servings per Container: about 14

Amount Per Serving	
Calories 110	Calories from Fat 15
	% Daily Value*
Total Fat 1.5g	2%

How much breadcrumbs does the recipe call for?

It calls for 1½ cups of breadcrumbs.

How many meatloaves does the recipe make?

It makes 1 meatloaf.

How many servings does that recipe make?

It says it serves 8.

How big is the canister?

It is cylindrical, 3.5 inches across and
7 inches tall.

What is the net weight of the contents of 1 canister?

15 ounces.

How much does a cup of breadcrumbs weigh?

I'm not sure, but maybe something from the nutritional label will help.

How much does a canister cost? $2.39

Info for bottle cap car

What kind of car is that?

A 1993 Honda Accord.

How big is that car / what are the dimensions? Here is some details from MSN autos:

Weight: 2800lb Length: 185.2 in Width: 67.1 in Height: 55.2 in

How much of the car was covered with caps?

Everything but the windows and the underside.

How big is a bottle cap?

Caps are 1 inch in diameter.

Info for chicken problem

How much chicken will you be buying?

Four pounds

How far are the two stores?

My neighborhood store is 2.2 miles away, and takes about 7 minutes. The store across town is 8.9 miles away, and takes about 25 minutes.

What kind of mileage does your car get?

It averages about 24 miles per gallon in the city.

How many gallons does your car hold?

About 14 gallons

How much is gas?

About $3.69/gallon right now.

Info for furnace problem

How efficient is the current furnace?

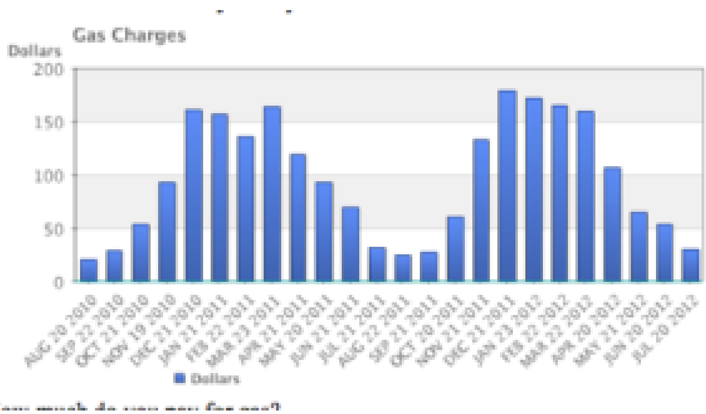
It is a 60% efficient furnace.

How efficient is the new furnace?

It is 94% efficient.

What is your gas bill?

Here is the history for 2 years:

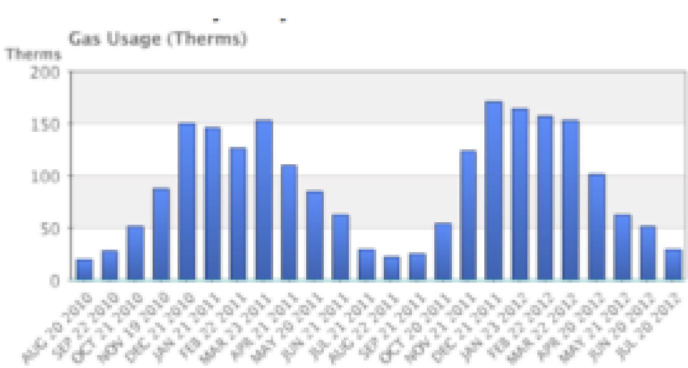


How much do you pay for gas?

There is $10.34 base charge, plus $0.39097 per Therm for a delivery charge, and $0.65195 per Therm for cost of gas.

How much gas do you use?

Here is the history for 2 years:



How much does the new furnace cost?

It will cost $7,450.

How long do you plan to live in the house?

Probably at least 15 years.

Info for water filter problem

How much water does Janine drink in a day?

She normally drinks 3 bottles a day, each 16.9 ounces.

How much does a bottle of water cost?

She buys 24-packs of 16.9 ounce bottles for $3.99.

How much does a reusable water bottle cost?

About $10.

How long does a reusable water bottle last?

Basically forever (or until you lose it).

How much does a water filter cost? How much water will they filter?

- A faucet-mounted filter costs about $28. Refill filters cost about $33 for a 3-pack. The box says each filter will filter up to 100 gallons (378 liters)
- A water filter pitcher costs about $22. Refill filters cost about $20 for a 4-pack. The box says each filter lasts for 40 gallons or 2 months
- An under-sink filter costs $130. Refill filters cost about

$60 each. The filter lasts for 500 gallons.

Info for car-free problem

Where does Marcus currently drive? He:

- Drives to work 5 days a week, located 4 miles from his house.
- Drives to the store twice a week, located 7 miles from his house.
- Drives to other locations on average 5 days a week, with locations ranging from 1 mile to 20 miles.
- Drives to his parent's house 80 miles away once a month.

How will he get to these locations without a car?

- For work, he can walk when it's sunny and he gets up early enough. Otherwise he can take a bus, which takes about 20 minutes
- For the store, he can take a bus, which takes about 35 minutes.
- Some of the other locations he can bus to. Sometimes he'll be able to get a friend to pick him up. A few locations he is able to walk to. A couple locations are hard to get to by bus, but there is a ZipCar (short term car rental) location within a few blocks.
- He'll need to get a ZipCar to visit his parents.

How much does gas cost?

About $3.69/gallon.

How much does he pay for insurance and maintenance?

- He pays $95/month for insurance.
- He pays $30 every 3 months for an oil change, and has averaged about $300/year for other maintenance costs.

How much is he paying for the car?

- He's paying $220/month on his car loan right now, and has 3 years left on the loan.
- If he sold the car, he'd be able to make enough to pay off the loan.
- If he keeps the car, he's planning on trading the car in for a newer model in a couple years.

What mileage does his car get?

About 26 miles per gallon on average.

How much does a bus ride cost?

$2.50 per trip, or $90 for an unlimited monthly pass.

How much does a ZipCar rental cost?

- The "occasional driving plan": $25 application fee and $60 annual fee, with no monthly commitment. Monday-Thursday the cost is $8/hour, or $72 per day. Friday-Sunday the cost is $8/hour or $78/day. Gas, insurance, and 180 miles are included in the cost. Additional miles are $0.45/mile.

- The "extra value plan": Same as above, but with a $50 monthly commitment, getting you a 10% discount on the usage costs.

[1] http://www.factcheck.org/cheney_edwards_mangle_facts.html

[2] http://tvtropes.org/pmwiki/pmwiki.php/Main/LiesDamnedLiesAndStatistics

[3] http://www.whitehouse.gov/sites/default/files/omb/budget/fy2013/assets/hist07z1.xls

[4] https://www.cia.gov/library/publications/the-world-factbook/rankorder/2042rank.html

[5] http://data.worldbank.org/indicator/SP.POP.TOTL

[6] http://www.cnn.com/2010/SHOWBIZ/Music/07/19/cd.digital.sales/index.html

[7] http://www.cnn.com/2012/06/27/politics/btn-health-care/index.html

[8] http://www.politico.com/news/stories/0712/78134.html

[9] http://factcheck.org/2012/07/twisting-health-care-taxes/

[10] http://www.northjersey.com/news/opinions/lautenberg_073112.html?c=y&page=2

[11] http://www.the-numbers.com/movies/records/budgets.php

[12] http://vimeo.com/42501010

[13] http://vimeo.com/40377128

[14] http://www.youtube.com/watch?v=DIkwefReHZc

[15] http://www.youtube.com/watch?v=p9SABH7Yg9M

[16] http://www.youtube.com/watch?v=62JMfv0tf3Q

[17] Photo credit: http://www.flickr.com/photos/swayze/, CC-BY

[18] http://www.notcot.com/archives/2009/06/floor-of-pennie.php

Attributions

CHAPTER II

CHAPTER 2: USING EXCEL

OVERVIEW OF MICROSOFT EXCEL

Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker

Learning Objectives

1. Examine the value of using Excel to make decisions.
2. Learn how to start Excel.
3. Become familiar with the Excel workbook.
4. Understand how to navigate worksheets.
5. Examine the Excel Ribbon.
6. Examine the right-click menu options.
7. Learn how to save workbooks.
8. Examine the Status Bar.
9. Become familiar with the features in the Excel Help window.

Microsoft® Office contains a variety of tools that help people accomplish many personal and professional objectives.

Microsoft Excel is perhaps the most versatile and widely used of all the Office applications. No matter which career path you choose, you will likely need to use Excel to accomplish your professional objectives, some of which may occur daily. This chapter provides an overview of the Excel application along with an orientation for accessing the commands and features of an Excel workbook. You will be using Excel in our work in statistics and math of finance.

Making Decisions with Excel

Taking a very simple view, Excel is a tool that allows you to enter quantitative data into an electronic spreadsheet to apply one or many mathematical computations. These computations ultimately convert that quantitative data into information. The information produced in Excel can be used to make decisions in both professional and personal contexts. For example, employees can use Excel to determine how much inventory to buy for a clothing retailer, how much medication to administer to a patient, or how much money to spend to stay within a budget. With respect to personal decisions, you can use Excel to determine how much money you can spend on a house, how much you can spend on car lease payments, or how much you need to save to reach your retirement goals. We will demonstrate how you can use Excel to make these decisions and many more throughout this text.

Figure 1.1 shows a completed Excel worksheet that will be constructed in this chapter. The information shown in this worksheet is top-line sales data for a hypothetical merchandise retail company. The worksheet data can help this retailer determine the number of salespeople needed for each month, how much inventory is needed to satisfy sales, and what types of products should be purchased.

A	B	C	D
General Merchandise World Retail Sales (in millions)			
Month	**Unit Sales**	**Average Price**	**Sales Dollars**
January	2,670	$ 9.99	$ 26,685
February	2,160	$ 12.49	$ 26,937
March	515	$ 14.99	$ 7,701
April	590	$ 17.49	$ 10,269
May	1,030	$ 14.99	$ 15,405
June	2,875	$ 12.49	$ 35,916
July	2,700	$ 9.99	$ 26,937
August	900	$ 19.99	$ 17,958
September	775	$ 19.99	$ 15,708
October	1,180	$ 19.99	$ 23,562
November	1,800	$ 17.49	$ 31,416
December	3,560	$ 14.99	$ 53,370
Total Sales	**20,755**		**$ 291,864**

Figure 1.1 Example of an Excel Worksheet

Starting Excel

1. Locate Excel on your computer.
2. Click Microsoft Excel to launch the Excel application and present you with workbook options.
3. Click the first option; “Blank Workbook”.

The Excel Workbook

Once Excel is started, a blank workbook will open on your screen. A workbook is an Excel file that contains one or more worksheets (sometimes referred to as spreadsheets). Excel will assign a file name to the workbook, such as **Book1**, **Book2**, **Book3**, and so on, depending on how many new workbooks are opened. **Figure 1.2** shows a blank workbook after starting Excel. Take some time to familiarize yourself with this screen. Your screen may be slightly different based on the version you’re using.

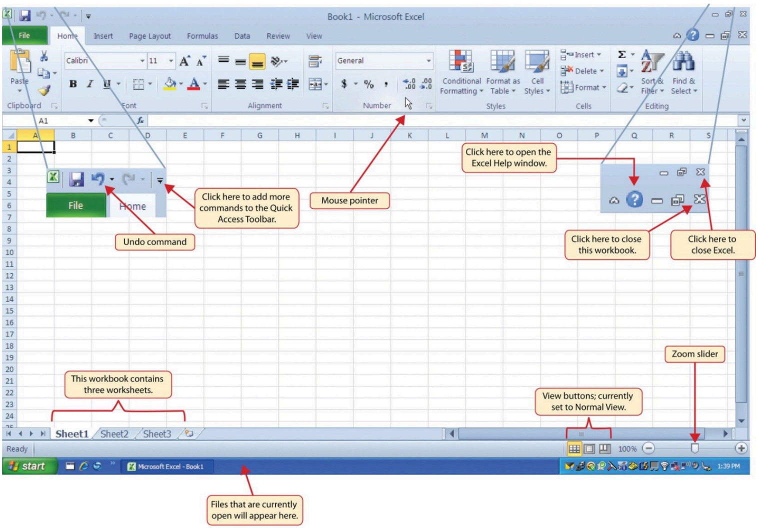


Figure 1.2 Blank Workbook

Your workbook should already be maximized (or shown at full size) once Excel is started, as shown in **Figure 1.2**. If necessary locate the Maximize button as shown in **Figure 1.3**.

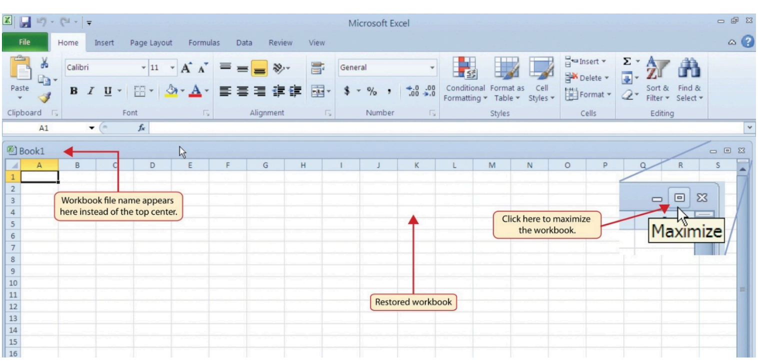


Figure 1.3 Restored Worksheet

Navigating Worksheets

Data are entered and managed in an Excel worksheet. The worksheet contains several rectangles called cells for entering numeric and nonnumeric data. Each cell in an Excel worksheet contains an address, which is defined by a column letter followed by a row number. For example, the cell that is currently activated in **Figure 1.3** is A1. This would be referred to as cell location A1 or cell reference A1. The following steps explain how you can navigate in an Excel worksheet:

1. Place your mouse pointer over cell D5 and left click.
2. Check to make sure column letter D and row number 5 are highlighted, as shown in **Figure 1.4**.

Note: Your highlighted column letter and row number may be different than figure shown.

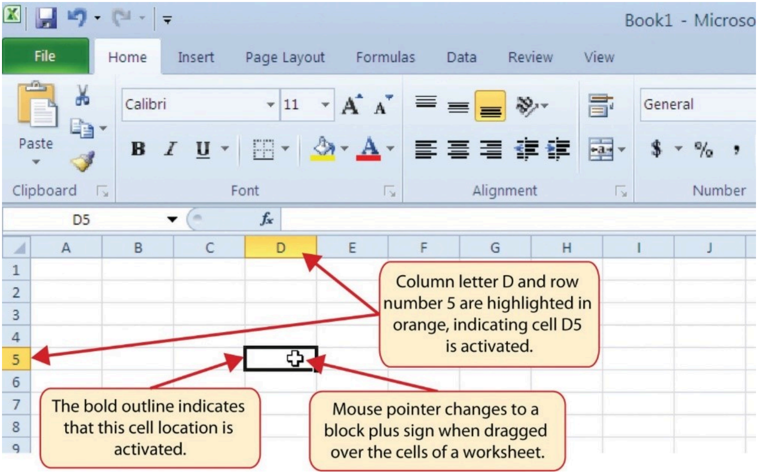


Figure 1.4 Activating a Cell Location

1. Move the mouse pointer to cell A1.
2. Click and hold the left mouse button and drag the mouse pointer back to cell D5.
3. Release the left mouse button. You should see several cells highlighted, as shown in **Figure 1.5**.

This is referred to as a *cell range* and is documented as follows: **A1:D5**. Any two cell locations separated by a colon are known as a cell range. The first cell is the top left corner of the range, and the second cell is the lower right corner of the range.

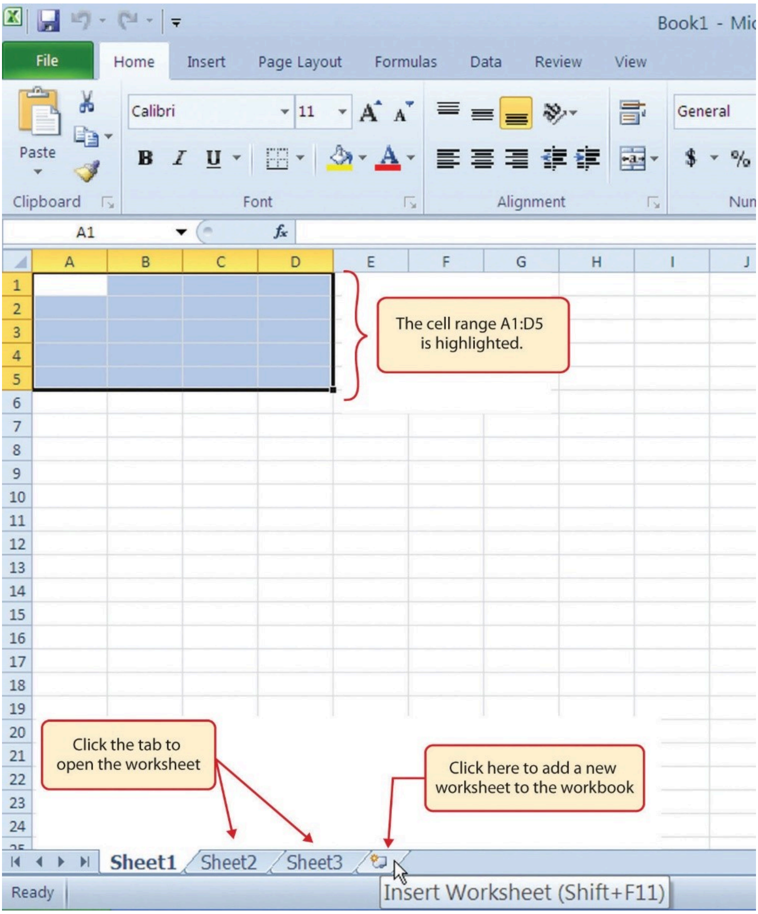


Figure 1.5 Highlighting a Range of Cells

1. At the bottom of the screen, you'll see worksheets. Depending on your version of Excel, you will see either three as displayed above or just one. If you only have one sheet, click the "Insert Worksheet" to add a worksheet. Depending on your version, you instead may have a +

sign; a click on the + adds an additional worksheet as well. This is how you open or add a worksheet within a workbook. Add another worksheet so that you now have three sheets displaying here.

2. Click the Sheet1 worksheet tab at the bottom of the worksheet to return to the worksheet shown in **Figure 1.5**.

Keyboard Shortcuts

Basic Worksheet Navigation

- Use the arrow keys on your keyboard to activate cells on the worksheet.
- Hold the SHIFT key and press the arrow keys on your keyboard to highlight a range of cells in a worksheet.
- Hold the CTRL key while pressing the PAGE DOWN or PAGE UP keys to open other worksheets in a workbook.

The Excel Ribbon

Excel's features and commands are found in the Ribbon, which is the upper area of the Excel screen that contains several tabs running across the top. Each tab provides access to a different set of Excel commands. **Figure 1.6** shows the commands available in the Home tab of the Ribbon. **Table 1.1 "Command Overview for Each Tab of the Ribbon"** provides an overview of the commands that are found in each tab of the Ribbon.

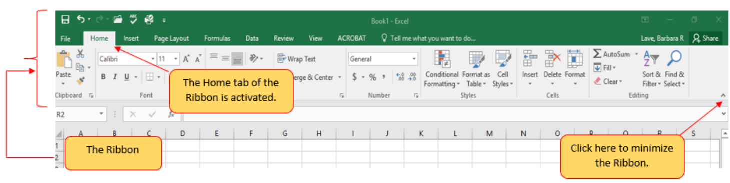


Figure 1.6 Home Tab of Ribbon

Table 1.1 Command Overview for Each Tab of the Ribbon

Tab Name	Description of Commands
File	Also known as the Backstage view of the Excel workbook. Contains all commands for opening, closing, saving, and creating new Excel workbooks. Includes print commands, document properties, e-mailing options, and help features. The default settings and options are also found in this tab.
Home	Contains the most frequently used Excel commands. Formatting commands are found in this tab along with commands for cutting, copying, pasting, and for inserting and deleting rows and columns.
Insert	Used to insert objects such as charts, pictures, shapes, PivotTables, Internet links, symbols, or text boxes.
Page Layout	Contains commands used to prepare a worksheet for printing. Also includes commands used to show and print the gridlines on a worksheet.
Formulas	Includes commands for adding mathematical functions to a worksheet. Also contains tools for auditing mathematical formulas.
Data	Used when working with external data sources such as Microsoft® Access®, text files, or the Internet. Also contains sorting commands and access to scenario tools.
Review	Includes Spelling and Track Changes features. Also contains protection features to password protect worksheets or workbooks.
View	Used to adjust the visual appearance of a workbook. Common commands include the Zoom and Page Layout view.

The Ribbon shown in **Figure 1.6** is full, or maximized. The benefit of having a full Ribbon is that the commands are always visible while you are developing a worksheet. However, depending on the screen dimensions of your computer, you may find that the Ribbon takes up too much vertical space on your worksheet. If this is the case, you can minimize the Ribbon by clicking the button shown in **Figure 1.6**. When minimized, the Ribbon will show only the tabs and not the command buttons. When you click on a tab, the command buttons will appear until you select a command or click anywhere on your worksheet.

Keyboard Shortcuts

Minimizing or Maximizing the Ribbon

- Hold down the CTRL key and press the F1 key.
- Hold down the CTRL key and press the F1 key again to maximize the Ribbon.

Quick Access Toolbar and Right-Click Menu

The Quick Access Toolbar is found at the upper left side of the Excel screen above the Ribbon, as shown in **Figure 1.7**. This area provides access to the most frequently used commands, such as Save and Undo. You also can customize the Quick Access Toolbar by adding commands that you use on a regular basis. By placing these commands in the Quick Access Toolbar, you do not have to navigate through the Ribbon to find them. To customize the Quick Access Toolbar, click the down arrow as shown in **Figure 1.7**. This will open a menu of commands that you can add to the Quick Access Toolbar. If you do not see the command you are looking for on the list, select the More Commands option.

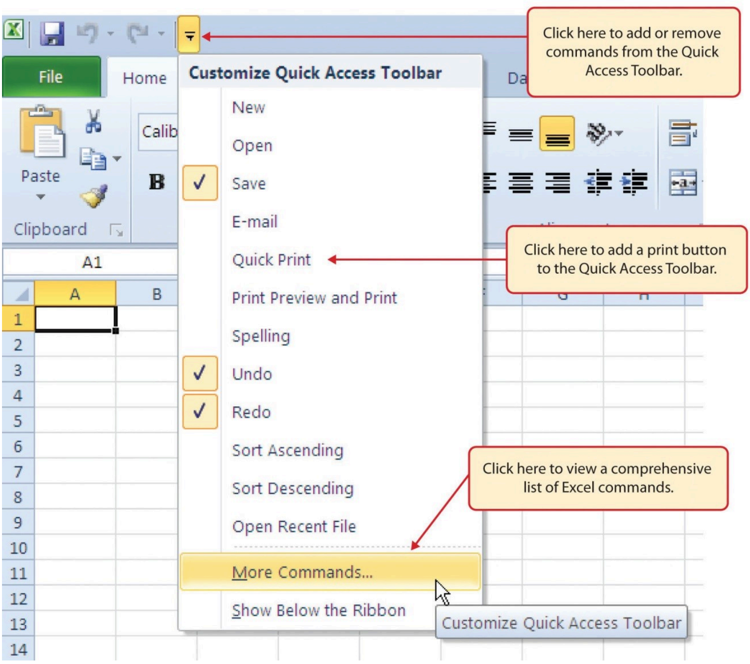


Figure 1.7 Customizing the Quick Access Toolbar

In addition to the Ribbon and Quick Access Toolbar, you can also access commands by right clicking anywhere on the worksheet. **Figure 1.8** shows an example of the commands available in the right-click menu.

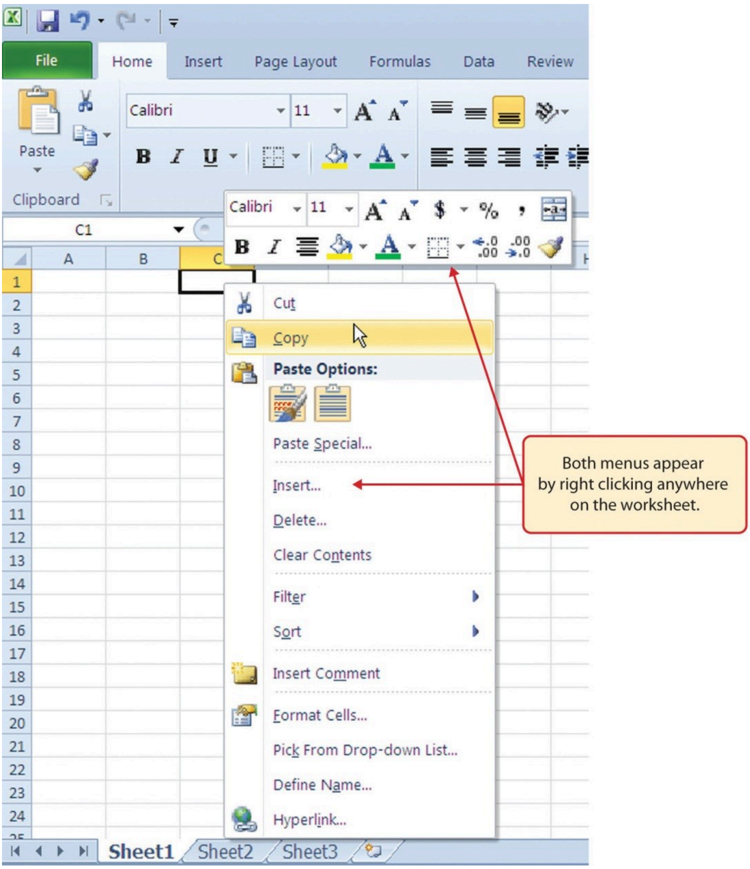


Figure 1.8 Right-Click Menu

The File Tab

The File tab is also known as the **Backstage** view of the workbook. It contains a variety of features and commands related to the workbook that is currently open, new workbooks, or workbooks stored in other locations on your computer or network. **Figure 1.9** shows the options available in the File tab or Backstage view. To leave the Backstage view and return to the worksheet, click the arrow in the upper left-hand corner as shown below.

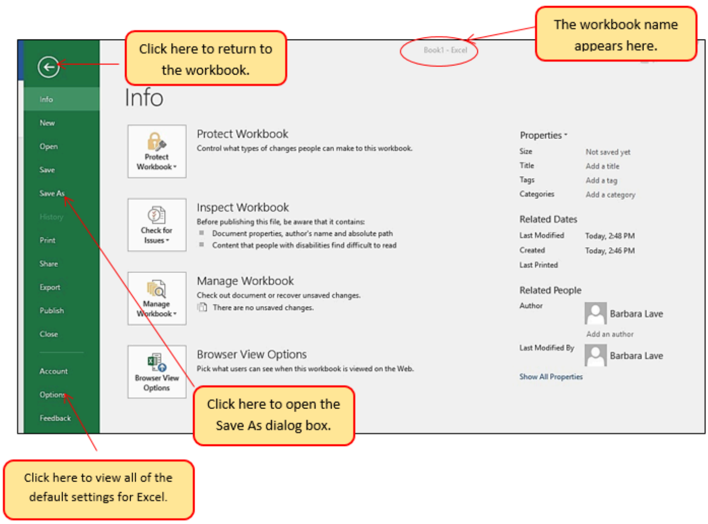


Figure 1.9 File Tab or Backstage View of a Workbook

Included in the File tab are the default settings for the Excel application that can be accessed and modified by clicking the Options button. **Figure 1.10** shows the Excel Options window, which gives you access to settings such as the default font style, font size, and the number of worksheets that appear in new workbooks.

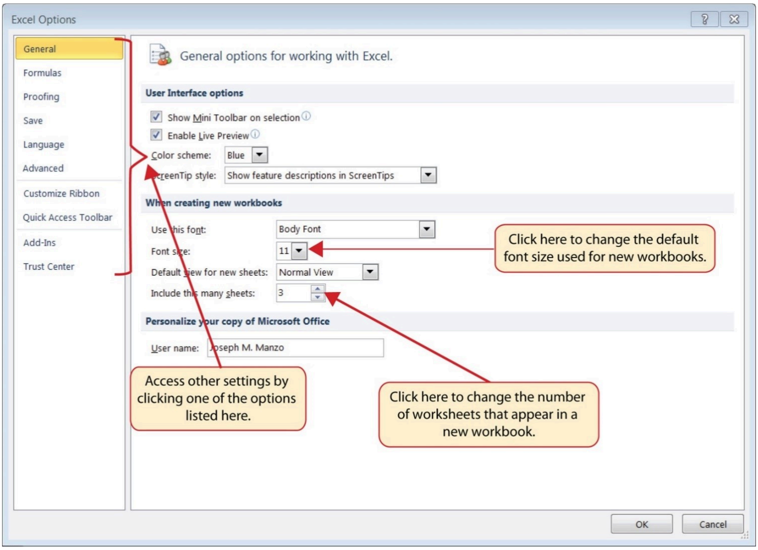


Figure 1.10 Excel Options Window

Saving Workbooks (Save As)

Once you create a new workbook, you will need to change the file name and choose a location on your computer or network to save that file. It is important to remember where you save this workbook on your computer or network as you will be using this file in the **Section 1.2 "Entering, Editing, and Managing Data"** to construct the workbook shown in **Figure 1.1**. The process of saving can be different with different versions of Excel. Please be sure you follow the steps for the version of Excel you are using. The following steps explain how to save a new workbook and assign it a file name.

Saving Workbooks in Excel 2013

1. If you have not done so already, open a blank workbook in Excel.
2. When saving your workbook for the *first* time, click the File tab.
3. Click the **Save As** button in the upper left side of the Backstage view window. This will open the **Save As** dialog box, as shown in **Figure 1.11**.
4. Click in the File Name box at the bottom of the **Save As** dialog box and use the BACKSPACE key to remove the current default name of the workbook.
5. Type the file name: **CH1 GMW Sales Data**.
6. Click the Desktop button on the left side of the **Save As** dialog box if you wish to save this file on your desktop. If you want to save this workbook in a different location, such as a USB drive, select your preferred location.
7. Click the Save button on the lower right side of the **Save As** dialog box.
8. As you continue to work on your workbook, you will want to Save frequently by click either the Save button on the Home ribbon; or by selecting the Save option from the File menu.

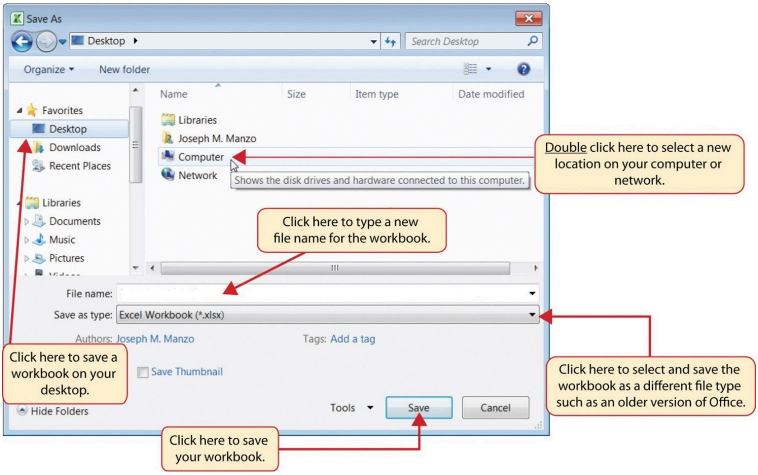


Figure 1.11 Save As Dialog Box in Excel 2013

Saving Workbooks in Excel 2016

1. If you have not done so already, open a blank workbook in Excel.
2. Click the File tab and then the **Save As** button in the left side of the Backstage view window. This will open the **Save As** dialog box.
3. Determine a location for saving on your computer by clicking **Browse** on the left side to open the **Save As** dialog box.
4. Click in the File Name box near the bottom of the **Save As** dialog box. Type the new file name: **CH1 GMW Sales Data**
5. Review the settings in the screen for correctness and

click the Save button.

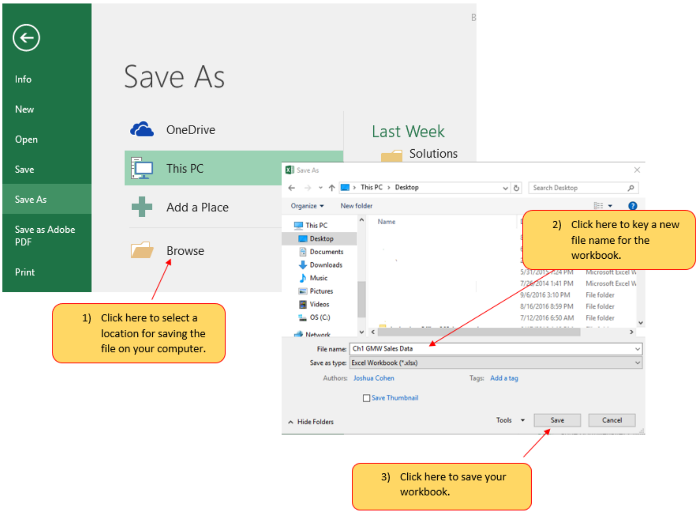


Figure 1.12 Save As Dialog in 2016

Keyboard Shortcuts

Save As

- Press the F12 key and use the tab and arrow keys to navigate around the Save As dialog box. Use the ENTER key to make a selection.
- Or press the ALT key on your keyboard. You

will see letters and numbers, called Key Tips, appear on the Ribbon. Press the F key on your keyboard for the File tab and then the A key. This will open the Save As dialog box.

Skill Refresher

Saving Workbooks (Save As)

1. Click the File tab on the Ribbon.
2. Click the Save As option.
3. Select a location on your PC.
4. Click in the File name box and type a new file name if needed.
5. Click the down arrow next to the “Save as type” box and select the appropriate file type if needed.
6. Click the Save button.

The Status Bar

The Status Bar is located below the worksheet tabs on the Excel screen (see Figure 1.13). It displays a variety of information, such as the status of certain keys on your keyboard (e.g., CAPS LOCK), the available views for a workbook, the magnification of the screen, and mathematical functions that can be performed when data are highlighted on a worksheet. You can customize the Status Bar as follows:

1. Place the mouse pointer over any area of the Status Bar and right click to display the "Customize Status Bar" list of options (**see Figure 1.13**).
2. Select the Caps Lock option from the menu (see **Figure 1.13**).
3. Press the CAPS LOCK key on your keyboard. You will see the Caps Lock indicator on the lower right side of the Status Bar.
4. Press the CAPS LOCK on your keyboard again. The indicator on the Status Bar goes away.

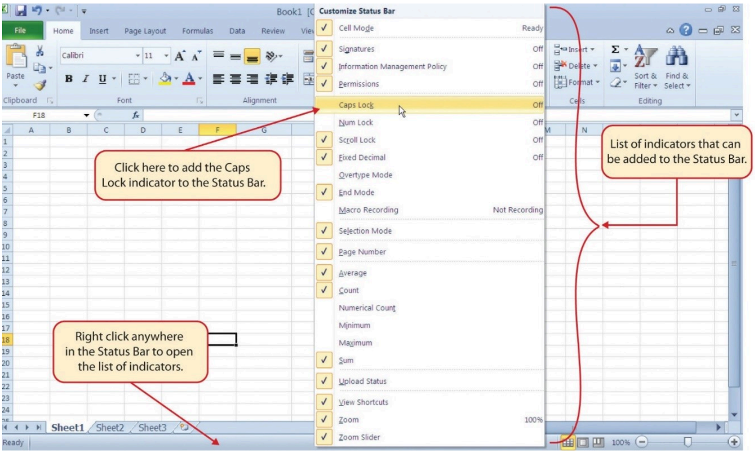


Figure 1.13 Customizing the Status Bar

Excel Help

The Help feature provides extensive information about the Excel application. Although some of this information may be stored on your computer, the Help window will automatically connect to the Internet, if you have a live connection, to provide you with resources that can answer most of your questions. You can open the Excel Help window by clicking the question mark in the upper right area of the screen or ribbon. With newer versions of Excel, use the query box to enter your question and select from helpful option links or select the question mark from the dropdown list to launch Excel Help windows.

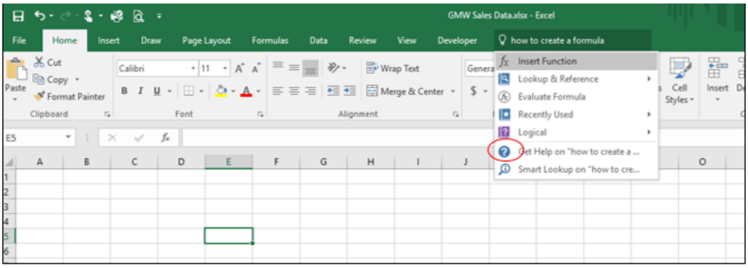


Figure 1.14 Excel Help Window

Keyboard Shortcuts

Excel Help

- Press the F1 key on your keyboard.

Key Takeaways

- Excel is a powerful tool for processing data for the purposes of making decisions.
- You can find Excel commands throughout the tabs in the Ribbon.
- You can customize the Quick Access Toolbar by

adding commands you frequently use.
- You can add or remove the information that is displayed on the Status Bar.
- The Help window provides you with extensive information about Excel.

Attribution

Adapted by Barbara Lave from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

Media Attributions

- figure_1-6_ribbon_for_excel
- figure-1-9
- figure_1-12_save_as_dialog_box_in_2016
- 1-14_excel_help_window

ENTERING, EDITING, AND MANAGING DATA

Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker

Learning Objectives

1. Understand how to enter data into a worksheet.
2. Examine how to edit data in a worksheet.
3. Examine how Auto Fill is used when entering data.
4. Understand how to delete data from a worksheet and use the Undo command.
5. Examine how to adjust column widths and row heights in a worksheet.
6. Understand how to hide columns and rows in a worksheet.
7. Examine how to insert columns and rows into a worksheet.
8. Understand how to delete columns and rows

from a worksheet.

9. Learn how to move data to different locations in a worksheet.

In this section, we will begin the development of the workbook shown in **Figure 1.1**. The skills covered in this section are typically used in the early stages of developing one or more worksheets in a workbook.

Entering Data

You will begin building the workbook shown in **Figure 1.1** by manually entering data into the worksheet. The following steps explain how the column headings in Row 2 are typed into the worksheet:

1. Click cell location A2 on the worksheet.
2. Type the word **Month**.
3. Press the RIGHT ARROW key. This will enter the word into cell A2 and activate the next cell to the right.
4. Type **Unit Sales** and press the RIGHT ARROW key.
5. Repeat step 4 for the words **Average Price** and then again for **Sales Dollars**.

Figure 1.15 shows how your worksheet should appear after you have typed the column headings into Row 2. Notice that the word **Price** in cell location C2 is not visible. This is because the column is too narrow to fit the entry you typed. We will examine formatting techniques to correct this problem in the next section.

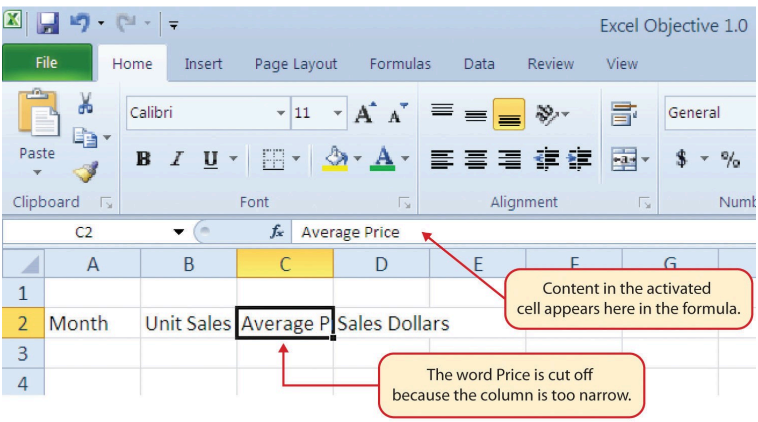


Figure 1.15 Entering Column Headings into a Worksheet

Integrity Check

Column Headings

It is critical to include column headings that accurately describe the data in each column of a worksheet. In professional environments, you will likely be sharing Excel workbooks with coworkers. Good column headings reduce the chance of someone misinterpreting the data contained in a worksheet, which could lead to costly errors depending on your career.

1. Click cell location B3.
2. Type the number **2670** and press the ENTER key. After you press the ENTER key, cell B4 will be activated. Using the ENTER key is an efficient way to enter data vertically down a column.
3. Enter the following numbers in cells B4 through B14: **2160**, **515, 590, 1030, 2875, 2700, 900, 775, 1 180, 1800**, and **3560**.
4. Click cell location C3.
5. Type the number **9.99** and press the ENTER key.
6. Enter the following numbers in cells C4 through C14: **12.49, 14.99, 17.49, 14.99, 12.49, 9.99, 19.99, 19.99, 19.99, 17.49**, and **14.99**.
7. Activate cell location D3.
8. Type the number 26685 and press the ENTER key.
9. Enter the following numbers in cells D4 through

D14: **26937, 7701, 10269, 15405, 35916, 26937, 17 958, 15708, 23562, 31416**, and **53370**.

10. When finished, check that the data you entered matches **Figure 1.16**.

Why?

Avoid Formatting Symbols When Entering Numbers

When typing numbers into an Excel worksheet, it is best to avoid adding any formatting symbols such as dollar signs and commas. Although Excel allows you to add these symbols while typing numbers, it slows down the process of entering data. It is more efficient to use Excel's formatting features to add these symbols to numbers after you type them into a worksheet.

Integrity Check

Data Entry

It is very important to proofread your worksheet carefully, especially when you have entered numbers. Transposing numbers when entering data manually into a worksheet is a common error. For example, the number **563** could be transposed to **536.** Such errors can seriously compromise the integrity of your workbook.

Integrity Check

Figure 1.16 shows how your worksheet should appear after entering the data. Check your numbers carefully to make sure they are accurately entered into the worksheet.

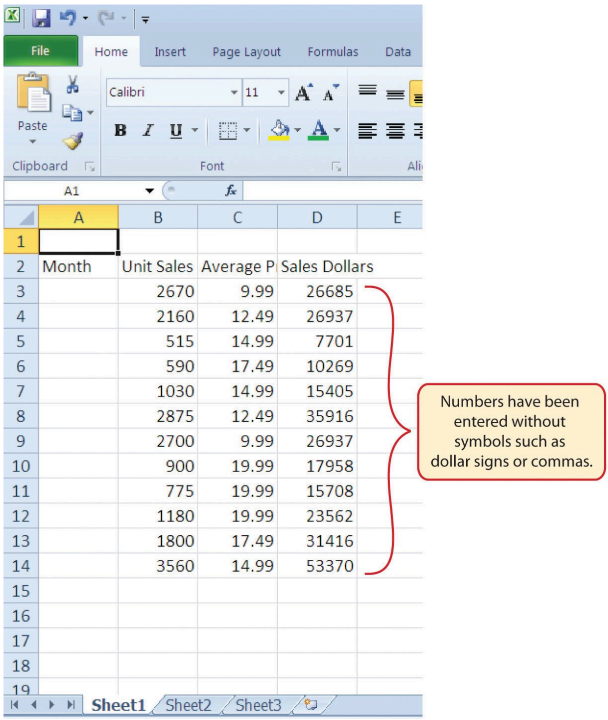


Figure 1.16 Completed Data Entry for Columns B, C, and D

Editing Data

Data that has been entered in a cell can be changed by double clicking the cell location or using the Formula Bar. You may have noticed that as you were typing data into a cell location, the data you typed appeared in the Formula Bar. The Formula Bar can be used for entering data into cells as well as for editing data that already exists in a cell. The following steps provide an example of entering and then editing data that has been entered into a cell location:

1. Click cell A15 in the Sheet1 worksheet.
2. Type the abbreviation **Tot** and press the ENTER key.
3. Click cell A15.
4. Move the mouse pointer up to the Formula Bar. You will see the pointer turn into a cursor. Move the cursor to the end of the abbreviation **Tot** and left click.
5. Type the letters **al** to complete the word Total.
6. Click the checkmark to the left of the Formula Bar (see **Figure 1.17**). This will enter the change into the cell.

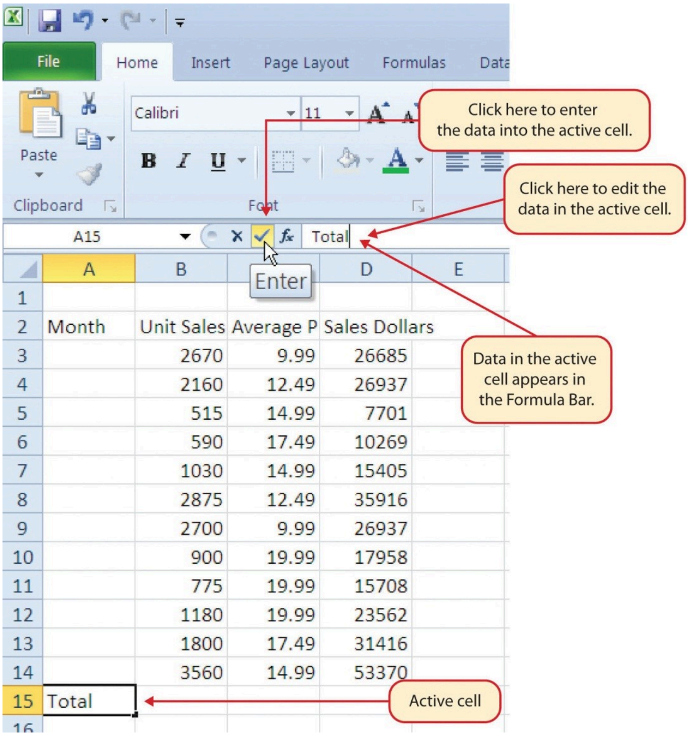


Figure 1.17 Using the Formula Bar to Edit and Enter Data

7. Double click cell A15.
8. Add a space after the word Total and type the word **Sales**.
9. Press the ENTER key.

Keyboard Shortcuts

Editing Data in a Cell

- Activate the cell that is to be edited and press the F2 key on your keyboard.

Auto Fill

The Auto Fill feature is a valuable tool when manually entering data into a worksheet. This feature has many uses, but it is most beneficial when you are entering data in a defined sequence, such as the numbers 2, 4, 6, 8, and so on, or nonnumeric data such as the days of the week or months of the year. The following steps demonstrate how Auto Fill can be used to enter the months of the year in Column A:

1. Click cell A3 in the Sheet1 worksheet.
2. Type the word **January** and press the ENTER key.
3. Activate cell A3 again.
4. Move the mouse pointer to the lower right corner of cell A3. You will see a small square in this corner of the cell; this is called the Fill Handle (See **Figure 1.18**) When the mouse pointer gets close to the Fill Handle, the white

block plus sign will turn into a black plus sign.

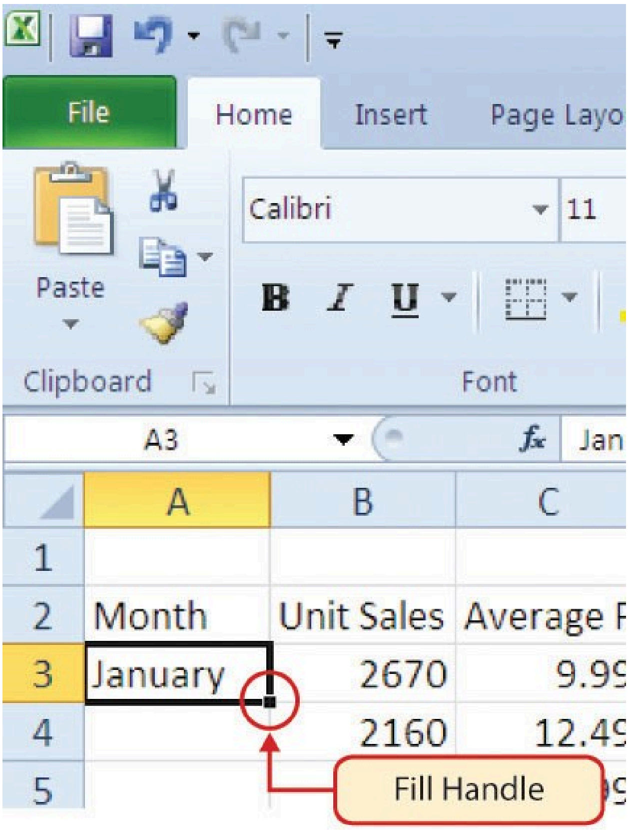


Figure 1.18 Fill Handle

Left click and drag the Fill Handle to cell A14. Notice that the Auto Fill tip box indicates what month will be placed into each cell (see **Figure 1.19**). Release the left mouse button when the tip box reads “December.”

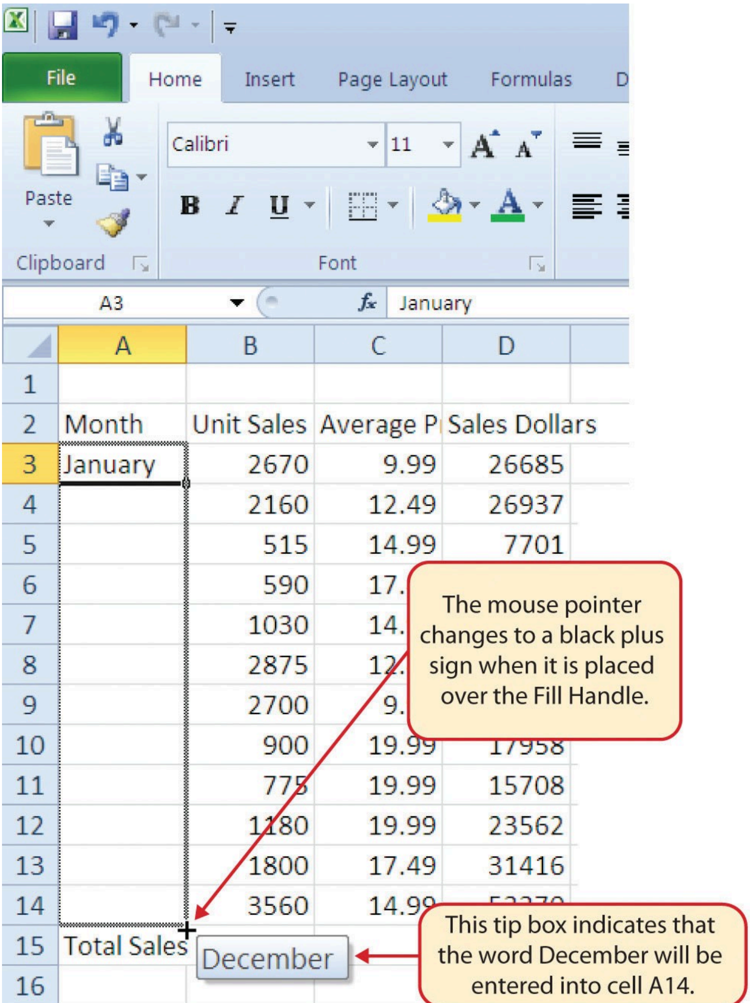


Figure 1.19 Using Auto Fill to Enter the Months of the Year

Once you release the left mouse button, all twelve months of the year should appear in the cell range A3:A14, as shown in **Figure 1.20**. You will also see the Auto Fill Options button.

By clicking this button, you have several options for inserting data into a group of cells.

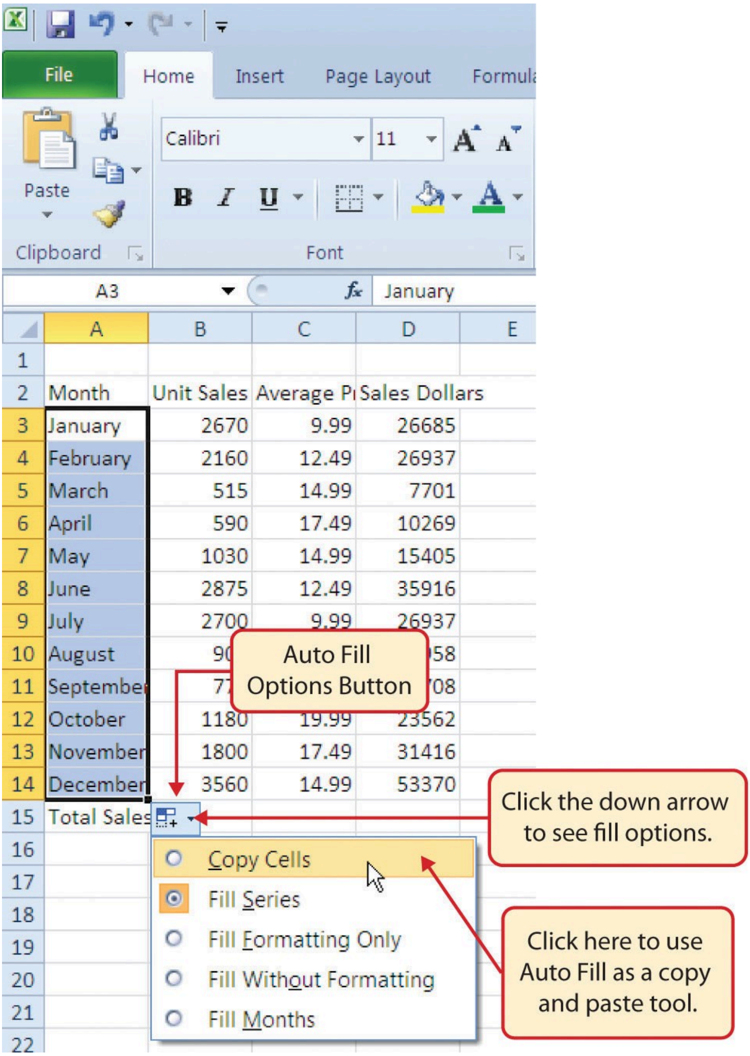


Figure 1.20 Auto Fill Options Button

1. Click the Auto Fill Options button.
2. Click the Copy Cells option. This will change the months in the range A4:A14 to January.
3. Click the Auto Fill Options button again.
4. Click the Fill Months option to return the months of the year to the cell range A4:A14. The Fill Series option will provide the same result.

Deleting Data and the Undo Command

There are several methods for removing data from a worksheet, a few of which are demonstrated here. With each method, you use the Undo command. This is a helpful command in the event you mistakenly remove data from your worksheet. The following steps demonstrate how you can delete data from a cell or range of cells:

1. Click cell C2 by placing the mouse pointer over the cell and clicking the left mouse button.
2. Press the **DELETE** key on your keyboard. This removes the contents of the cell.
3. Highlight the range C3:C14 by placing the mouse pointer over cell C3. Then left click and drag the mouse pointer down to cell C14.
4. Place the mouse pointer over the Fill Handle. You will see the white block plus sign change to a black plus sign.

5. Click and drag the mouse pointer up to cell C3 (see **Figure 1.21**). Release the mouse button. The contents in the range C3:C14 will be removed.

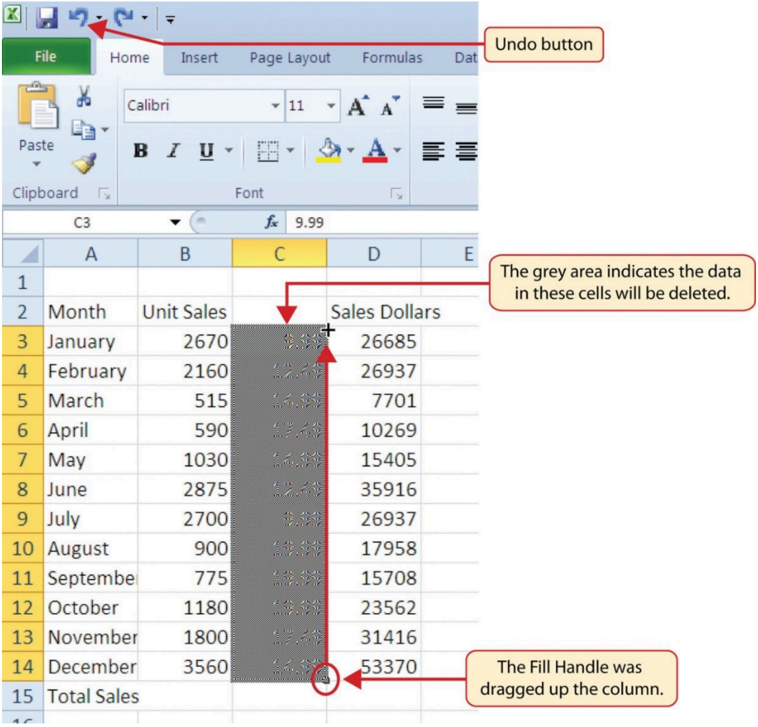


Figure 1.21 Using Auto Fill to Delete Contents of Cell

1. Click the Undo button in the Quick Access Toolbar (see **Figure 1.2**). This should replace the data in the range C3:C14.
2. Click the Undo button again. This should replace the data in cell C2.

Keyboard Shortcuts

Undo Command

- Hold down the CTRL key while pressing the letter Z on your keyboard.
- Highlight the range C2:C14 by placing the mouse pointer over cell C2. Then left click and drag the mouse pointer down to cell C14.
- Click the Clear button in the Home tab of the Ribbon, which is next to the Cells group of commands (see **Figure 1.22**). This opens a drop-down menu that contains several options for removing or clearing data from a cell. Notice that you also have options for clearing just the formats in a cell or the hyperlinks in a cell.
- Click the Clear All option. This removes the data in the cell range.
- Click the Undo button. This replaces the data in the range C2:C14.

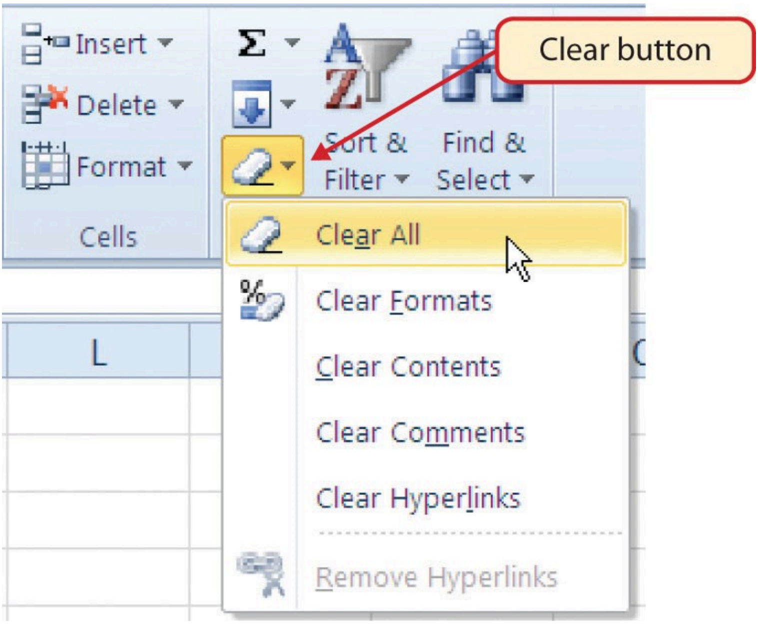


Figure 1.22 Clear Command Drop-Down Menu

Adjusting Columns and Rows

There are a few entries in the worksheet that appear cut off. For example, the last letter of the word September cannot be seen in cell A11. This is because the column is too narrow for this word. The columns and rows on an Excel worksheet can be adjusted to accommodate the data that is being entered into a cell. The following steps explain how to adjust the column widths and row heights in a worksheet:

1. Bring the mouse pointer between Column A and

Column B in the Sheet1 worksheet, as shown in **Figure 1.23**. You will see the white block plus sign turn into double arrows.

2. Click and drag the column to the right so the entire word September in cell A11 can be seen. As you drag the column, you will see the column width tip box. This box displays the number of characters that will fit into the column using the Calibri 11-point font which is the default setting for font/size.
3. Release the left mouse button.

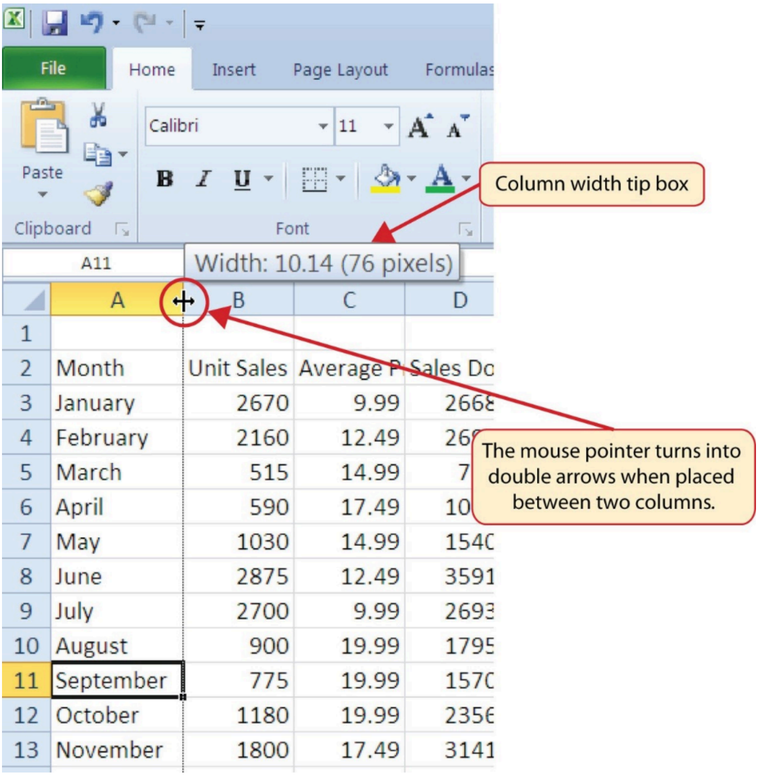


Figure 1.23 Adjusting Column Widths

You may find that using the click-and-drag method is inefficient if you need to set a specific character width for one or more columns. Steps 1 through 6 illustrate a second method for adjusting column widths when using a specific number of characters:

1. Click any cell location in Column A by moving the mouse pointer over a cell location and clicking the left

mouse button. You can highlight cell locations in multiple columns if you are setting the same character width for more than one column.

2. In the Home tab of the Ribbon, left click the Format button in the Cells group.
3. Click the Column Width option from the drop-down menu. This will open the Column Width dialog box.
4. Type the number **13** and click the OK button on the Column Width dialog box. This will set Column A to this character width (see **Figure 1.24**).
5. Once again bring the mouse pointer between Column A and Column B so that the double arrow pointer displays and then double-click to activate AutoFit. This features adjusts the column width based on the longest entry in the column.
6. Use the Column Width dialog box (step 6 above) to reset the width to 13.

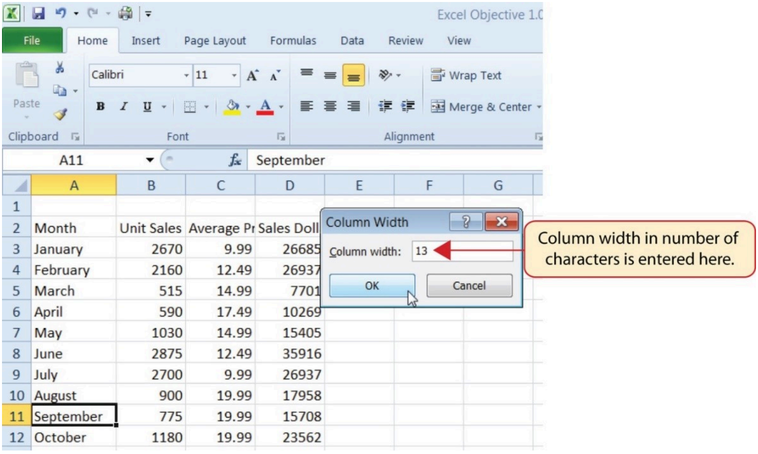


Figure 1.24 Column Width Dialog Box

Keyboard Shortcuts

Column Width

- Press the ALT key on your keyboard, then press the letters H, O, and W one at a time.

Steps 1 through 4 demonstrate how to adjust row height, which is similar to adjusting column width:

1. Click cell A15 by placing the mouse pointer over the cell and clicking the left mouse button.

2. In the Home tab of the Ribbon, left click the Format button in the Cells group.
3. Click the Row Height option from the drop-down menu. This will open the Row Height dialog box.
4. Type the number **24** and click the OK button on the Row Height dialog box. This will set Row 15 to a height of 24 points. A point is equivalent to approximately 1/72 of an inch. This adjustment in row height was made to create space between the totals for this worksheet and the rest of the data.

Keyboard Shortcuts

Row Height

- Press the ALT key on your keyboard, then press the letters H, O, and H one at a time.

Figure 1.25 shows the appearance of the worksheet after Column A and Row 15 are adjusted.

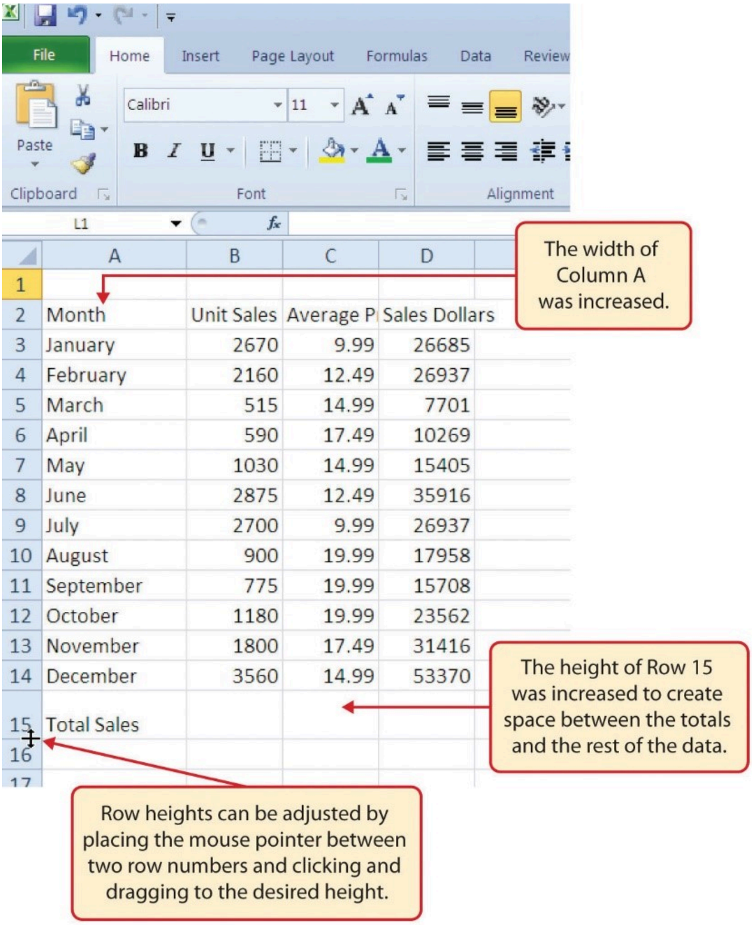


Figure 1.25 GMW Sales Data with Column A and Row 15 Adjusted

Skill Refresher

Adjusting Columns and Rows

1. Activate at least one cell in the row or column you are adjusting.
2. Click the Home tab of the Ribbon.
3. Click the Format button in the Cells group.
4. Click either Row Height or Column Width from the drop-down menu.
5. Enter the Row Height in points or Column Width in characters in the dialog box.
6. Click the OK button.

Hiding Columns and Rows

In addition to adjusting the columns and rows on a worksheet, you can also hide columns and rows. This is a useful technique for enhancing the visual appearance of a worksheet that contains data that is not necessary to display. These features will be demonstrated using the GMW Sales Data workbook. However, there is no need to have hidden columns or rows for this worksheet. The use of these skills here will be for demonstration purposes only.

1. Click cell C1 in the Sheet1 worksheet by placing the mouse pointer over the cell location and clicking the left mouse button.
2. Click the Format button in the Home tab of the Ribbon.
3. Place the mouse pointer over the Hide & Unhide option in the drop-down menu. This will open a submenu of options.
4. Click the Hide Columns option in the submenu of options (see **Figure 1.26**). This will hide Column C.

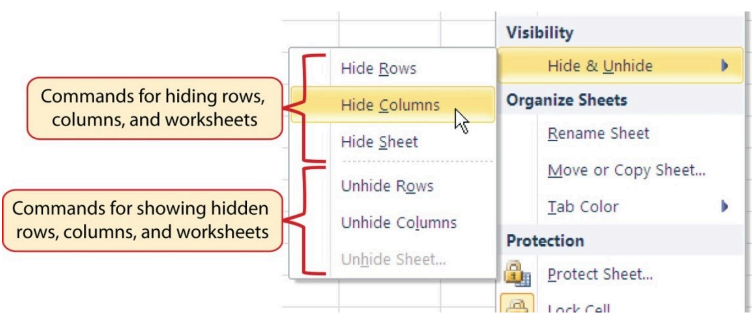


Figure 1.26 Hide & Unhide Submenu

Keyboard Shortcuts

Hiding Columns

- Hold down the CTRL key while pressing the number 0 on your keyboard.

Figure 1.27 shows the workbook with Column C hidden in the Sheet1 worksheet. You can tell a column is hidden by the missing letter C.

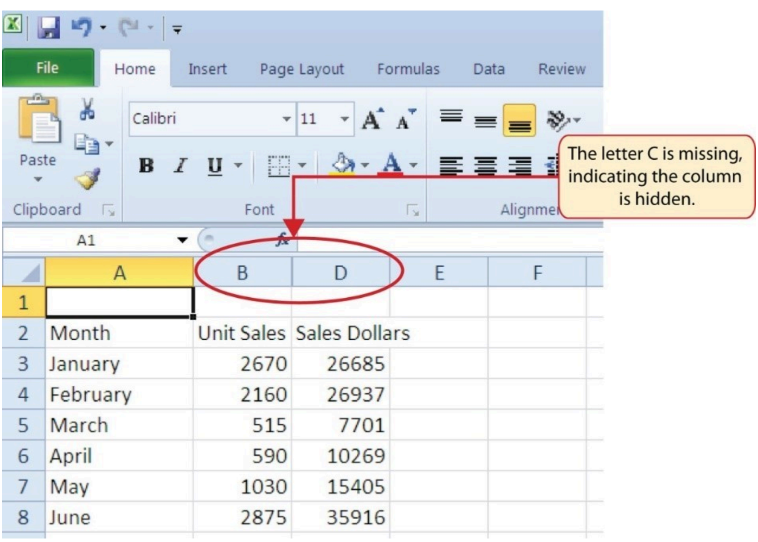


Figure 1.27 Hidden Column

To unhide a column, follow these steps:

1. Highlight the range B1:D1 by activating cell B1 and clicking and dragging over to cell D1.
2. Click the Format button in the Home tab of the Ribbon.
3. Place the mouse pointer over the Hide & Unhide option in the drop-down menu.
4. Click the Unhide Columns option in the submenu of options. Column C will now be visible on the worksheet.

Keyboard Shortcuts

Unhiding Columns

- Highlight cells on either side of the hidden column(s), then hold down the CTRL key and the SHIFT key while pressing the close parenthesis key ()) on your keyboard.

The following steps demonstrate how to hide rows, which is similar to hiding columns:

1. Click cell A3 in the Sheet1 worksheet by placing the

mouse pointer over the cell location and clicking the left mouse button.

2. Click the Format button in the Home tab of the Ribbon.
3. Place the mouse pointer over the Hide & Unhide option in the drop-down menu. This will open a submenu of options.
4. Click the Hide Rows option in the submenu of options. This will hide Row 3.

Keyboard Shortcuts

Hiding Rows

- Hold down the CTRL key while pressing the number 9 key on your keyboard.

To unhide a row, follow these steps:

1. Highlight the range A2:A4 by activating cell A2 and clicking and dragging over to cell A4.
2. Click the Format button in the Home tab of the Ribbon.
3. Place the mouse pointer over the Hide & Unhide option in the drop-down menu.

4. Click the Unhide Rows option in the submenu of options. Row 3 will now be visible on the worksheet.

Keyboard Shortcuts

Unhiding Rows

- Highlight cells above and below the hidden row(s), then hold down the CTRL key and the SHIFT key while pressing the open parenthesis key (() on your keyboard.

Integrity Check

Hidden Rows and Columns

In most careers, it is common for professionals to use Excel workbooks that have been designed by a coworker. Before you use a workbook developed by someone else, always check for hidden rows and columns. You can quickly see whether a row or column is hidden if a row number or column letter is missing.

Skill Refresher

Hiding Columns and Rows

1. Activate at least one cell in the row(s) or column(s) you are hiding.
2. Click the Home tab of the Ribbon.
3. Click the Format button in the Cells group.
4. Place the mouse pointer over the Hide & Unhide option.
5. Click either the Hide Rows or Hide Columns option.

Skill Refresher

Unhiding Columns and Rows

1. Highlight the cells above and below the hidden row(s) or to the left and right of the hidden column(s).
2. Click the Home tab of the Ribbon.
3. Click the Format button in the Cells group.
4. Place the mouse pointer over the Hide & Unhide option.
5. Click either the Unhide Rows or Unhide Columns option.

Inserting Columns and Rows

Using Excel workbooks that have been created by others is a very efficient way to work because it eliminates the need to create data worksheets from scratch. However, you may find that to accomplish your goals, you need to add additional columns or rows of data. In this case, you can insert blank columns or rows into a worksheet. The following steps demonstrate how to do this:

1. Click cell C1 in the Sheet1 worksheet by placing the mouse pointer over the cell location and clicking the left mouse button.
2. Click the down arrow on the Insert button in the Home tab of the Ribbon (see **Figure 1.28**).

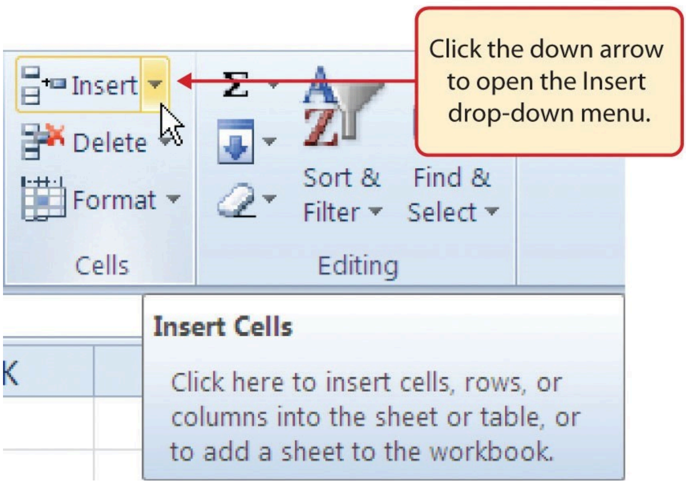


Figure 1.28 Insert Button (Down Arrow)

3. Click the Insert Sheet Columns option from the drop-down menu (see **Figure 1.29**). A blank column will be inserted to the left of Column C. The contents that were previously in Column C now appear in Column D. Note that columns are always inserted to the left of the activated cell.

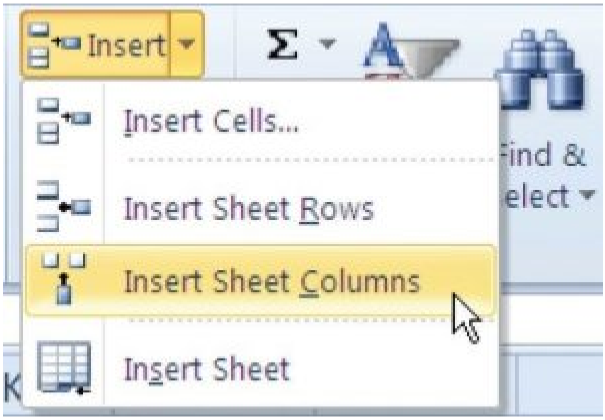


Figure 1.29 Insert Drop-Down Menu

Keyboard Shortcuts

Inserting Columns

- Press the ALT key and then the letters H, I, and C one at a time. A column will be inserted to the left of the activated cell.

4. Click cell A3 in the **Sheet1** worksheet by placing the mouse pointer over the cell location and clicking the left mouse button.

5. Click the down arrow on the Insert button in the Home tab of the Ribbon (see **Figure 1.28**).
6. Click the Insert Sheet Rows option from the drop-down menu (see **Figure 1.29**). A blank row will be inserted above Row 3. The contents that were previously in Row 3 now appear in Row 4. Note that rows are always inserted above the activated cell.

Keyboard Shortcuts

Inserting Rows

- Press the ALT key and then the letters H, I, and R one at a time. A row will be inserted above the activated cell.

Skill Refresher

Inserting Columns and Rows

1. Activate the cell to the right of the desired blank column or below the desired blank row.

2. Click the Home tab of the Ribbon.
3. Click the down arrow on the Insert button in the Cells group.
4. Click either the Insert Sheet Columns or Insert Sheet Rows option.

Moving Data

Once data are entered into a worksheet, you have the ability to move it to different locations. The following steps demonstrate how to move data to different locations on a worksheet:

1. Highlight the range D2:D15 by activating cell D2 and clicking and dragging down to cell D15.
2. Bring the mouse pointer to the left edge of cell D2. You will see the white block plus sign change to cross arrows (see **Figure 1.30**). This indicates that you can left click and drag the data to a new location.

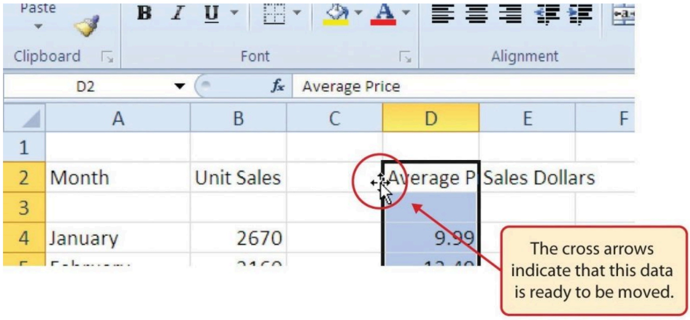


Figure 1.30 Moving Data

3. Left Click and drag the mouse pointer to cell C2.
4. Release the left mouse button. The data now appears in Column C.
5. Click the Undo button in the Quick Access Toolbar. This moves the data back to Column D.

Integrity Check

Moving Data

Before moving data on a worksheet, make sure you identify all the components that belong with the series you are moving. For example, if you are moving a column of data, make sure the column heading is included. Also, make sure all values are highlighted in the column before moving it.

Deleting Columns and Rows

You may need to delete entire columns or rows of data from a worksheet. This need may arise if you need to remove either blank columns or rows from a worksheet or columns and rows that contain data. The methods for removing cell contents were covered earlier and can be used to delete unwanted data. However, if you do not want a blank row or column in your workbook, you can delete it using the following steps:

1. Click cell A3 by placing the mouse pointer over the cell location and clicking the left mouse button.
2. Click the down arrow on the Delete button in the Cells group in the Home tab of the Ribbon.
3. Click the Delete Sheet Rows option from the drop-down menu (see **Figure 1.31**). This removes Row 3 and shifts all the data (below Row 2) in the worksheet up one row.

Keyboard Shortcuts

Deleting Rows

- Press the ALT key and then the letters H, D, and R one at a time. The row with the activated cell will be deleted.

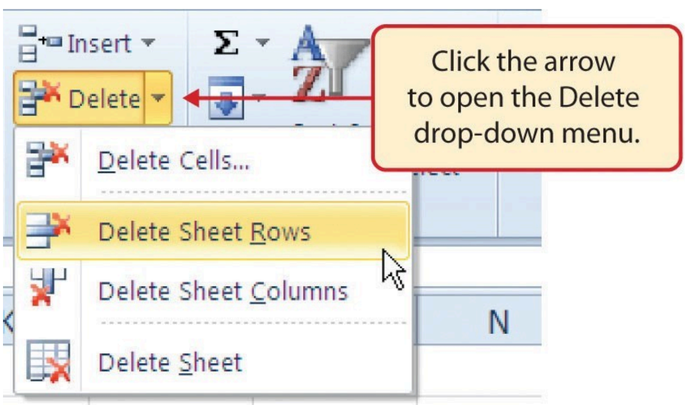


Figure 1.31 Delete Drop-Down Menu

4. Click cell C1 by placing the mouse pointer over the cell location and clicking the left mouse button.
5. Click the down arrow on the Delete button in the Cells group in the Home tab of the Ribbon.

6. Click the Delete Sheet Columns option from the drop-down menu (see **Figure 1.31**). This removes Column C and shifts all the data in the worksheet (to the right of Column B) over one column to the left.
7. Save the changes to your workbook by clicking either the **Save** button on the Home ribbon; or by selecting the **Save** option from the File menu.

Keyboard Shortcuts

Deleting Columns

- Press the ALT key and then the letters H, D, and C one at a time. The column with the activated cell will be deleted.

Skill Refresher

Deleting Columns and Rows

1. Activate any cell in the row or column that is to be deleted.

2. Click the Home tab of the Ribbon.
3. Click the down arrow on the Delete button in the Cells group.
4. Click either the Delete Sheet Columns or the Delete Sheet Rows option.

Key Takeaways

- Column headings should be used in a worksheet and should accurately describe the data contained in each column.
- Using symbols such as dollar signs when entering numbers into a worksheet can slow down the data entry process.
- Worksheets must be carefully proofread when data has been manually entered.
- The Undo command is a valuable tool for recovering data that was deleted from a worksheet.
- When using a worksheet that was developed by someone else, look carefully for hidden columns or rows.

Attribution

Media Attributions

- 1-2-fig-1-27
- 1-2-fig-1-28
- 1-2-fig-1-29
- 1-2-fig-1-30
- 1-2-fig-1-31

FORMATTING AND DATA ANALYSIS

Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker

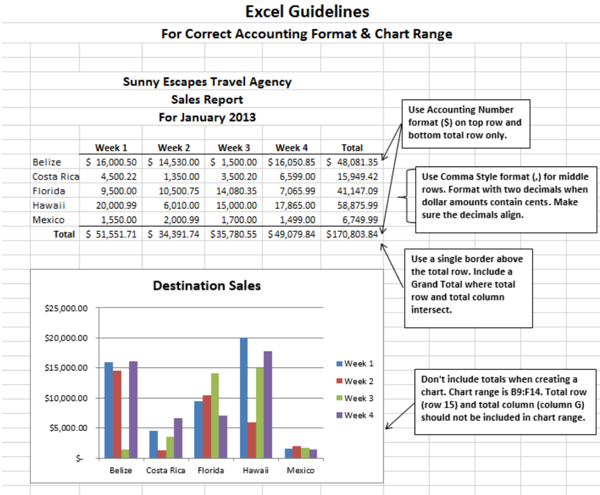
Learning Objectives

1. Use formatting techniques as introduced in the Excel Spreadsheet Guidelines to enhance the appearance of a worksheet.
2. Understand how to align data in cell locations.
3. Examine how to enter multiple lines of text in a cell location.
4. Understand how to add borders to a worksheet.
5. Examine how to use the AutoSum feature to calculate totals.
6. Use the Cut, Copy, and Paste commands to manipulate the data on a worksheet.
7. Understand how to move, rename, insert, and delete worksheet tabs.

This section addresses formatting commands that can be used to enhance the visual appearance of a worksheet. It also provides an introduction to mathematical calculations. The skills introduced in this section will give you powerful tools for analyzing the data that we have been working with in this workbook and will highlight how Excel is used to make key decisions in virtually any career. Additionally, Excel Spreadsheet Guidelines for format and appearance will be introduced as a format for the course and spreadsheets submitted.

Formatting Data and Cells

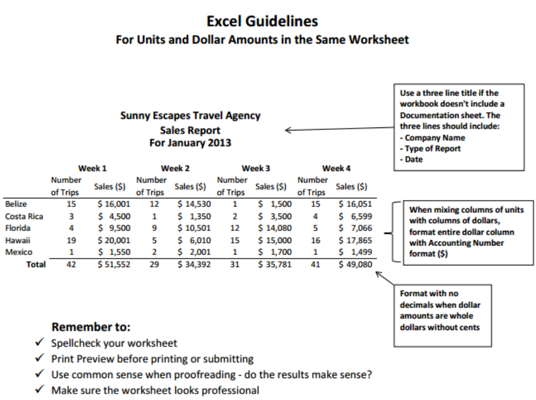
Enhancing the visual appearance of a worksheet is a critical step in creating a valuable tool for you or your coworkers when making key decisions. There are accepted professional formatting standards when spreadsheets contain only currency data. For this course, we will use the following Excel Guidelines for Formatting. The first figure displays how to use Accounting number format when ALL figures are currency. Only the first row of data and the totals should be formatted with the Accounting format. The other data should be formatted with Comma style. There also needs to be a Top Border above the numbers in the total row. If any of the numbers have cents, you need to format all of the data with two decimal places.



	Week 1	Week 2	Week 3	Week 4	Total
Belize	$ 16,000.50	$ 14,530.00	$ 1,500.00	$ 16,050.85	$ 48,081.35
Costa Rica	4,500.22	1,350.00	3,500.20	6,599.00	15,949.42
Florida	9,500.00	10,500.75	14,080.35	7,065.99	41,147.09
Hawaii	20,000.99	6,010.00	15,000.00	17,865.00	58,875.99
Mexico	1,550.00	2,000.99	1,700.00	1,499.00	6,749.99
Total	$ 51,551.71	$ 34,391.74	$35,780.55	$49,079.84	$170,803.84

Figure 1.31a

Often, your Excel spreadsheet will contain values that are both currency and non-currency in nature. When that is the case, you'll want to use the guidelines in the following figure:



	Week 1		Week 2		Week 3		Week 4	
	Number of Trips	Sales ($)	Number of Trips	Sales ($)	Number of Trips	Sales ($)	Number of Trips	Sales ($)
Belize	15	$ 16,001	12	$ 14,530	1	$ 1,500	15	$ 16,051
Costa Rica	3	$ 4,500	1	$ 1,350	2	$ 3,500	4	$ 6,599
Florida	4	$ 9,500	9	$ 10,501	12	$ 14,080	5	$ 7,066
Hawaii	19	$ 20,001	5	$ 6,010	15	$ 15,000	16	$ 17,865
Mexico	1	$ 1,550	2	$ 2,001	1	$ 1,700	1	$ 1,499
Total	42	$ 51,552	29	$ 34,392	31	$ 35,781	41	$ 49,080

Figure 1.31b

The following steps demonstrate several fundamental formatting skills that will be applied to the workbook that we are developing for this chapter. Several of these formatting skills are identical to ones that you may have already used in other Microsoft applications such as Microsoft® Word® or Microsoft® PowerPoint®.

1. Highlight the range A2:D2 in the **Sheet1** worksheet by placing the mouse pointer over cell A2 and left clicking and dragging to cell D2. Click the Bold button in the Font group of commands in the Home tab of the ribbon.
2. Click the Border button in the Font group of commands in the Home tab of the Ribbon (see **Figure 1.32**). Select the Bottom Border option from the list to achieve the goal of a border on the bottom of row 2 below the column headings.

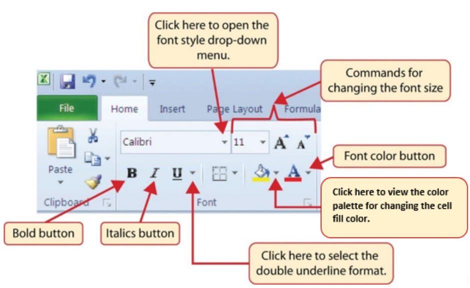


Figure 1.32 Font Group of Commands

Keyboard Shortcuts

Bold Format

- Hold the CTRL key while pressing the letter B on your keyboard.

3. Highlight the range A15:D15 by placing the mouse pointer over cell A15 and left clicking and dragging to cell D15.
4. Click the Bold button in the Font group of commands in the Home tab of the Ribbon.
5. Click the Border button in the Font group of commands in the Home tab of the Ribbon (see **Figure 1.32**). Select the Top Border option from the list to achieve the goal of a border on the top of row 15 where totals will eventually display.

Keyboard Shortcuts

Italics Format

- Hold the CTRL key while pressing the letter I on your keyboard.

Keyboard Shortcuts

Underline Format

- Hold the CTRL key while pressing the letter U on your keyboard.

Why?

Format Column Headings and Totals

Applying formatting enhancements to the column headings and column totals in a worksheet is a very important technique, especially if you are sharing a workbook with other people. These formatting techniques allow users of the worksheet to clearly see the column headings that define the data. In addition, the column totals usually contain the most important data on a worksheet with respect to making decisions, and formatting techniques allow users to quickly see this information.

1. Highlight the range B3:B14 by placing the mouse pointer over cell B3 and left clicking and dragging down to cell B14.
2. Click the Comma Style button in the Number group of commands in the Home tab of the Ribbon. This feature adds a comma as well as two decimal places. (see **Figure 1.33**).

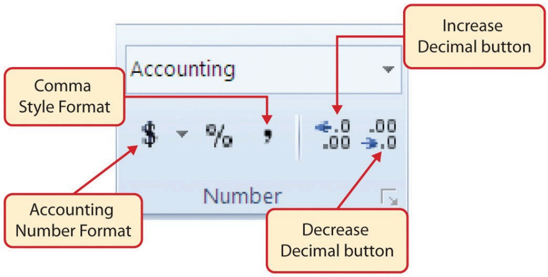


Figure 1.33 Number Group of Commands

3. Since the figures in this range do not include cents, click the Decrease Decimal button in the Number group of commands in the Home tab of the Ribbon two times (see **Figure 1.33**).
4. The numbers will also be reduced to zero decimal places.
5. Highlight the range C3:C14 by placing the mouse pointer over cell C3 and left clicking and dragging down to cell C14.
6. Click the Accounting Number Format button in the Number group of commands in the Home tab of the Ribbon (see **Figure 1.33**). This will add the US currency symbol and two decimal places to the values. This format is common when working with pricing data. As discussed above in the Formatting Data and Cells section, you will want to use Accounting format on all values in this range since the worksheet contains non-currency as well as currency data.
7. Highlight the range D3:D14 by placing the mouse pointer over cell D3 and left clicking and dragging down to cell D14.

8. Again, select the Accounting Number Format; this will add the US currency symbol to the values as well as two decimal places.
9. Click the Decrease Decimal button in the Number group of commands in the Home tab of the Ribbon.
10. This will add the US currency symbol to the values and reduce the decimal places to zero since there are no cents in these figures.
11. Highlight the range A1:D1 by placing the mouse pointer over cell A1 and left clicking and dragging over to cell D1.
12. Click the down arrow next to the Fill Color button in the Font group of commands in the Home tab of the Ribbon (see **Figure 1.34**). This will prepare the range for a worksheet title.

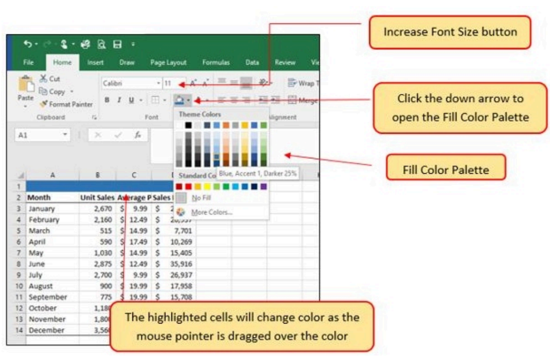


Figure 1.34 Fill Color Palette

13. Click the Blue, Accent 1, Darker 25% color from the palette (see **Figure 1.34**). Notice that as you move the

mouse pointer over the color palette, you will see a preview of how the color will appear in the highlighted cells. Experiment with this feature.

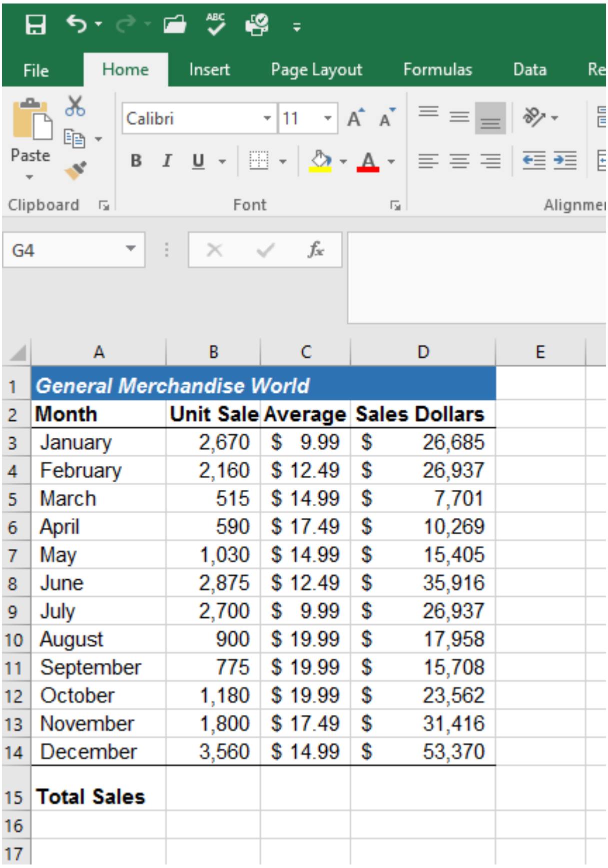
14. Click on A1 and enter the worksheet title: General Merchandise World and click on the check mark in the formula bar to enter this information.
15. Since the black font is difficult to read on the blue background, you'll change the font color to be more visible. Click the down arrow next to the Font Color button in the Font group of commands in the Home tab of the Ribbon; select White as the font color for this range (see **Figure 1.32**).
16. Highlight the range A1:D15 by placing the mouse pointer over cell A1 and left clicking and dragging down to cell D15.
17. Click the drop-down arrow on the right side of the Font button in the Home tab of the Ribbon; select Arial as the font for this range. (see **Figure 1.32**).
18. Notice that as you move the mouse pointer over the font style options, you can see the font change in the highlighted cells.
19. Expand the column width of Column D to 14 characters.

Why?

Pound Signs (####) Appear in Columns

When a column is too narrow for a long number, Excel will automatically convert the number to a series of pound signs (####). In the case of words or text data, Excel will only show the characters that fit in the column. However, this is not the case with numeric data because it can give the appearance of a number that is much smaller than what is actually in the cell. To remove the pound signs, increase the width of the column.

Figure 1.35 shows how the Sheet1 worksheet should appear after the formatting techniques are applied.



	A	B	C	D
1	General Merchandise World			
2	Month	Unit Sale	Average	Sales Dollars
3	January	2,670	$ 9.99	$ 26,685
4	February	2,160	$ 12.49	$ 26,937
5	March	515	$ 14.99	$ 7,701
6	April	590	$ 17.49	$ 10,269
7	May	1,030	$ 14.99	$ 15,405
8	June	2,875	$ 12.49	$ 35,916
9	July	2,700	$ 9.99	$ 26,937
10	August	900	$ 19.99	$ 17,958
11	September	775	$ 19.99	$ 15,708
12	October	1,180	$ 19.99	$ 23,562
13	November	1,800	$ 17.49	$ 31,416
14	December	3,560	$ 14.99	$ 53,370
15	Total Sales			

Figure 1.35 Formatting Techniques Applied

Data Alignment (Wrap Text, Merge Cells, and Center)

The skills presented in this segment show how data are aligned within cell locations. For example, text and numbers can be centered in a cell location, left justified, right justified, and so on. In some cases you may want to stack multiword text entries vertically in a cell instead of expanding the width of a column. This is referred to as wrapping text. These skills are demonstrated in the following steps:

1. Highlight the range B2:D2 by placing the mouse pointer over cell B2 and left clicking and dragging over to cell D2.
2. Click the Center button in the Alignment group of commands in the Home tab of the Ribbon (see **Figure 1.36**). This will center the column headings in each cell location.

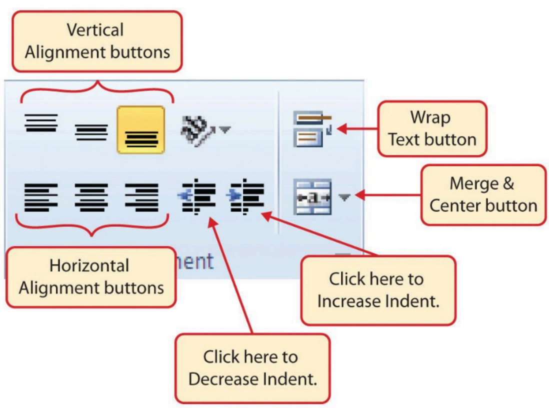


Figure 1.36 Alignment Group in Home Tab

3. Click the Wrap Text button in the Alignment group (see **Figure 1.36**). The height of Row 2 automatically expands, and the words that were cut off because the columns were too narrow are now stacked vertically.

Keyboard Shortcuts

Wrap Text

- Press the ALT key and then the letters H and W one at a time.

Why?

Wrap Text

The benefit of using the Wrap Text command is that it significantly reduces the need to expand the column width to accommodate multiword column headings. The problem with increasing the column width is that you may reduce the amount of data that can fit on a piece of paper or one screen. This makes it cumbersome to analyze the data in the worksheet and could increase the time it takes to make a decision.

4. Highlight the range A1:D1 by placing the mouse pointer over cell A1 and left clicking and dragging over to cell D1.
5. Click the down arrow on the right side of the Merge & Center button in the Alignment group of commands in the Home tab of the Ribbon.
6. Left click the Merge & Center option (see **Figure 1.37**). This will create one large cell location running across the top of the data set.

Keyboard Shortcuts

Merge Commands

- Merge & Center: Press the ALT key and then the letters H, M, and C one at a time.
- Merge Cells: Press the ALT key and then the letters H, M, and M one at a time.
- Unmerge Cells: Press the ALT key and then the letters H, M, and U one at a time.

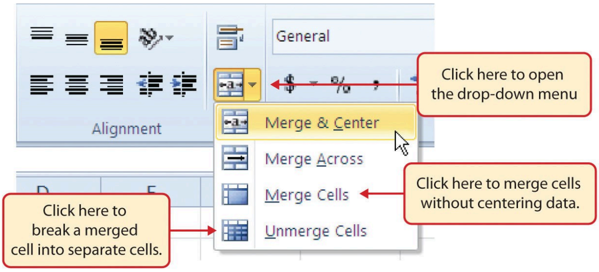


Figure 1.37 Merge Cell Drop-Down Menu

Why?

Merge & Center

One of the most common reasons the Merge & Center command is used is to center the title of a worksheet directly above the columns of data. Once the cells above the column headings are merged, a title can be centered above the columns of data. It is very difficult to center the title over the columns of data if the cells are not merged.

Figure 1.38 shows the Sheet1 worksheet with the data alignment commands applied. The reason for merging the cells in the range A1:D1 will become apparent in the next segment.

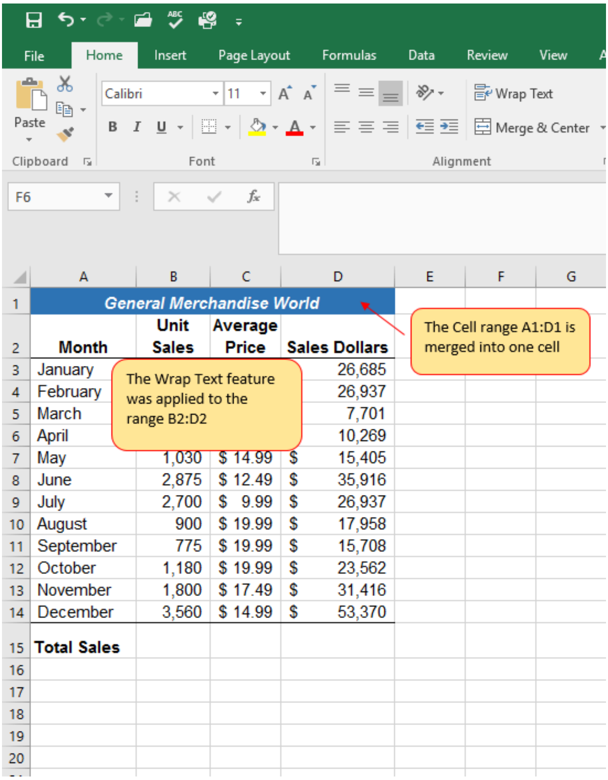


Figure 1.38 Sheet1 with Data Alignment Features Added

Skill Refresher

Wrap Text

1. Activate the cell or range of cells that contain

text data.
2. Click the Home tab of the Ribbon.
3. Click the Wrap Text button.

Skill Refresher

Merge Cells

1. Highlight a range of cells that will be merged.
2. Click the Home tab of the Ribbon.
3. Click the down arrow next to the Merge & Center button.
4. Select an option from the Merge & Center list.

Entering Multiple Lines of Text

In the Sheet1 worksheet, the cells in the range A1:D1 were merged for the purposes of adding a title to the worksheet. This worksheet will contain both a title and a subtitle. The following steps explain how you can enter text into a cell and determine where you want the second line of text to begin:

1. Activate cell A1 in the Sheet1 worksheet by placing the mouse pointer over cell A1 and clicking the left mouse button. Since the cells were merged, clicking cell A1 will automatically activate the range A1:D1. Position your mouse to the end of the title, directly after the "d" in the word "World" and double-click to get a cursor (flashing I-beam).
2. Hold down the ALT key and press the ENTER key. This will start a new line of text in this cell location.
3. Type the text **Retail Sales (in millions)** and press the ENTER key.
4. Select cell A1. Then click the Italics and Bold buttons in the Font group of commands in the Home tab of the Ribbon.
5. Increase the height of Row 1 to 30 points. Once the row height is increased, all the text typed into the cell will be visible (see **Figure 1.39**).

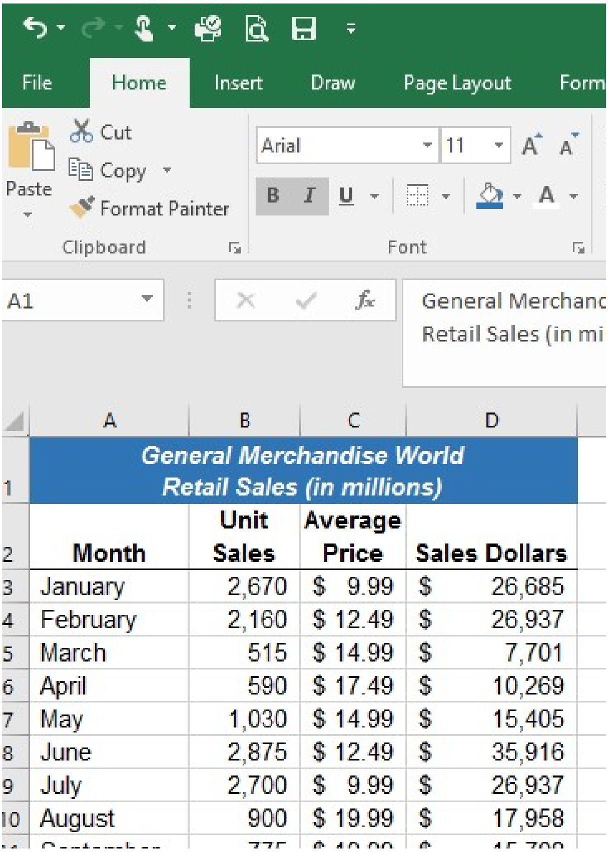


Figure 1.39 Title & Subtitle Added to the Worksheet

Skill Refresher

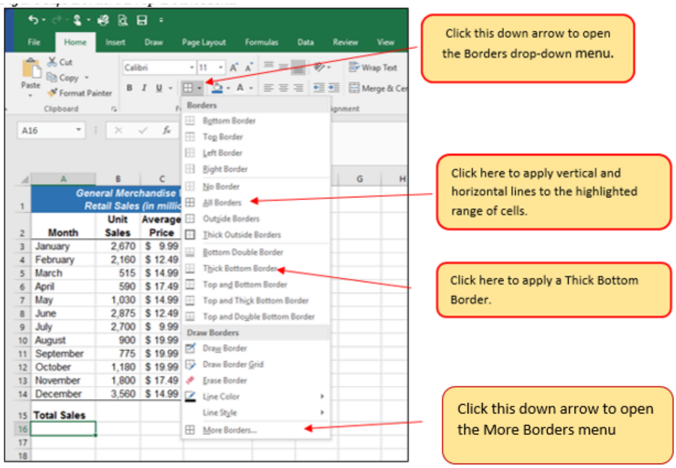
Entering Multiple Lines of Text

1. Activate a cell location.
2. Type the first line of text.
3. Hold down the ALT key and press the ENTER key.
4. Type the second line of text and press the ENTER key.

Borders (Adding Lines to a Worksheet)

In Excel, adding custom lines to a worksheet is known as adding borders. Borders are different from the grid lines that appear on a worksheet and that define the perimeter of the cell locations. The Borders command lets you add a variety of line styles to a worksheet that can make reading the worksheet much easier. The following steps illustrate methods for adding preset borders and custom borders to a worksheet:

1. Click the down arrow to the right of the Borders button in the Font group of commands in the Home page of the Ribbon to view border options. (see **Figure 1.40**).Figure 1.40 Borders Drop-Down Menu



2. Highlight the range A1:D15. Left click the All Borders option from the Borders drop-down menu (see **Figure 1.40**). This will add vertical and horizontal lines to the range A1:D15.
3. Highlight the range A2:D2 by placing the mouse pointer over cell A2 and left clicking and dragging over to cell D2.
4. Click the down arrow to the right of the Borders button.
5. Left click the Thick Bottom Border option from the Borders drop-down menu.
6. Highlight the range A14:D14 and apply a Thick Bottom Border from the drop-down menu. The thick border will help maintain the Excel Formatting Guidelines.
7. Highlight the range A1:D15.
8. Click the down arrow to the right of the Borders button.
9. Click More Borders… at the bottom of the List.
10. This will open the Format Cells dialog box (see **Figure**

1.41). You can access all formatting commands in Excel through this dialog box.

11. In the Style section of the Borders tab, left click the thickest line style (see **Figure 1.41**).
12. Left click the Outline button in the Presets section (see **Figure 1.41**).
13. Click the OK button at the bottom of the dialog box (see **Figure 1.41**).

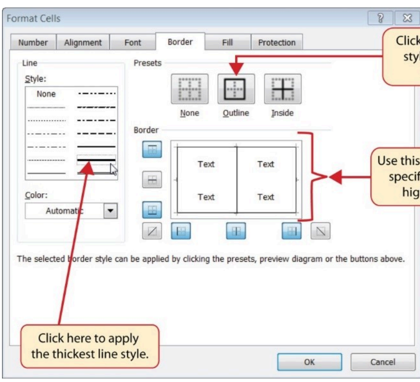


Figure 1.41 Borders Tab of the Format Cells Dialog Box

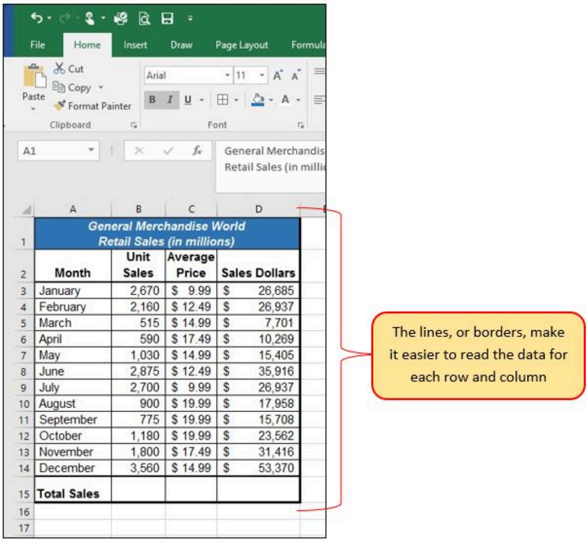


Figure 1.42 Borders Added to the Sheet1 Worksheet

Skill Refresher

Preset Borders

1. Highlight a range of cells that require borders.
2. Click the Home tab of the Ribbon.
3. Click the down arrow next to the Borders button.
4. Select an option from the preset borders list.

Custom Borders

1. Highlight a range of cells that require borders.
2. Click the Home tab of the Ribbon.
3. Click the down arrow next to the Borders button.
4. Select the More Borders option at the bottom of the options list.
5. Select a line style and line color.
6. Select a placement option.
7. Click the OK button on the dialog box.

AutoSum

You will see at the bottom of Figure 1.42 that Row 15 is intended to show the totals for the data in this worksheet. Applying mathematical computations to a range of cells is accomplished through functions in Excel. Chapter 2 “Mathematical Computations” will review mathematical formulas and functions in detail. However, the following steps will demonstrate how you can quickly sum the values in a column of data using the AutoSum command:

1. Activate cell B15 in the Sheet1 worksheet.

2. Click the Formulas tab of the Ribbon.
3. Click the down arrow below the AutoSum button in the Function Library group of commands (see **Figure 1.43**). Note that the AutoSum button can also be found in the Editing group of commands in the Home tab of the Ribbon.

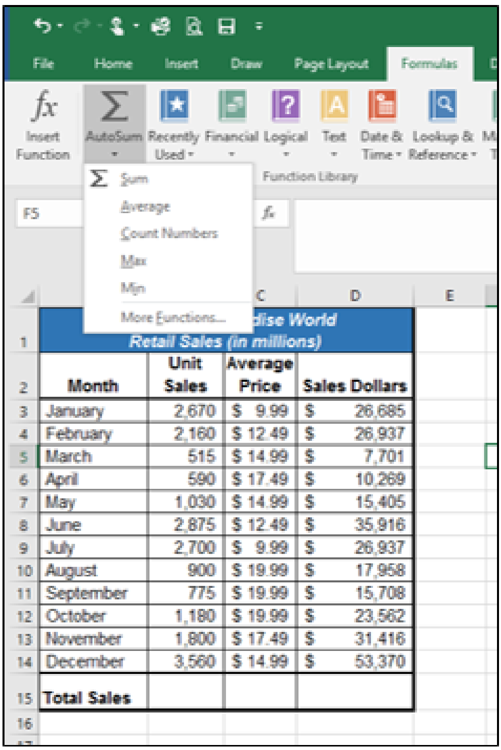
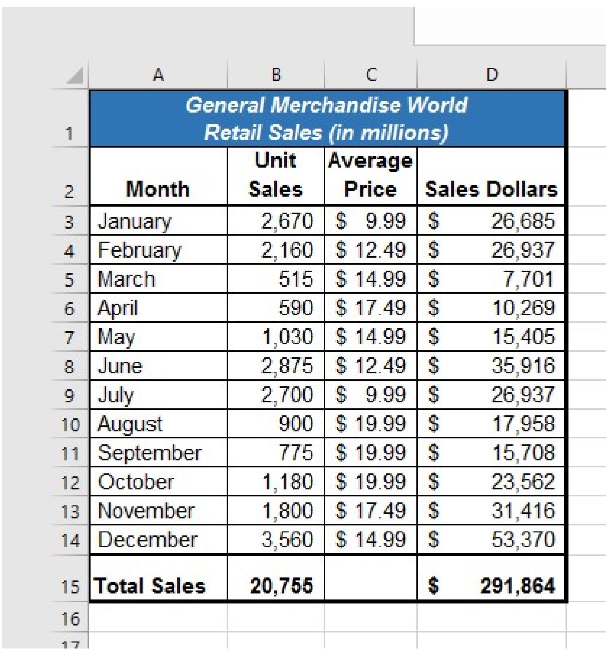


Figure 1.43 AutoSum Drop-Down List

4. Click the Sum option from the AutoSum drop-down menu. The first click will display a flashing marquee

around the range. Click the check mark next to the Formula bar to complete the function.

5. Excel will provide a total for the values in the Unit Sales column.
6. Activate cell D15. It would not make sense to total the averages in column C so C15 will be left blank.
7. Repeat steps 3 through 5 to sum the values in the Sales Dollars column (see **Figure 1.44**).



	A	B	C	D
1	*General Merchandise World Retail Sales (in millions)*			
2	**Month**	**Unit Sales**	**Average Price**	**Sales Dollars**
3	January	2,670	$ 9.99	$ 26,685
4	February	2,160	$ 12.49	$ 26,937
5	March	515	$ 14.99	$ 7,701
6	April	590	$ 17.49	$ 10,269
7	May	1,030	$ 14.99	$ 15,405
8	June	2,875	$ 12.49	$ 35,916
9	July	2,700	$ 9.99	$ 26,937
10	August	900	$ 19.99	$ 17,958
11	September	775	$ 19.99	$ 15,708
12	October	1,180	$ 19.99	$ 23,562
13	November	1,800	$ 17.49	$ 31,416
14	December	3,560	$ 14.99	$ 53,370
15	**Total Sales**	**20,755**		**$ 291,864**
16				

Figure 1.44 Totals Added to the Sheet1 Worksheet

Skill Refresher

AutoSum

1. Highlight a cell location below or to the right of a range of cells that contain numeric values.
2. Click the Formulas tab of the Ribbon.
3. Click the down arrow below the AutoSum button.
4. Select a mathematical function from the list.

Moving, Renaming, Inserting, and Deleting Worksheets

The default names for the worksheet tabs at the bottom of workbook are Sheet1, Sheet2, and so on. However, you can change the worksheet tab names to identify the data you are using in a workbook. Additionally, you can change the order in which the worksheet tabs appear in the workbook. The following steps explain how to rename and move the worksheets in a workbook:

1. With the left mouse button, double click the Sheet1 worksheet tab at the bottom of the

workbook (see **Figure 1.45**). Type the name **Sales by Month**.

2. Press the ENTER key on your keyboard.
3. With the left mouse button, double click the Sheet2 worksheet tab at the bottom of the workbook.
4. Type the name Unit Sales Rank to prepare the worksheet for future use.
5. Press the ENTER key on your keyboard.

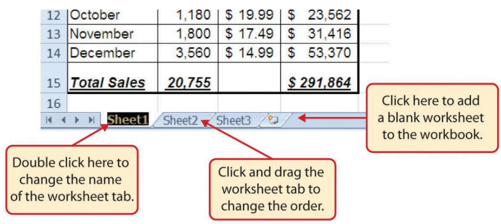


Figure 1.45 Renaming a Worksheet Tab

6. Click the Sheet3 worksheet tab.
7. Click the Home tab of the Ribbon.
8. Click the down arrow on the Delete button in the Cells group of commands.
9. Click the Delete Sheet option from the drop-down list. This removes the unneeded worksheet.
10. Click the Delete button on the Delete warning box (if a warning box appears).
11. Complete the steps above to delete the newly

named Unit Sales Rank worksheet since it's decided that worksheet is also unnecessary so that you are left with just one worksheet.

12. Save the changes to your workbook by clicking either the **Save** button on the Home ribbon; or by selecting the **Save** option from the File menu.

Integrity Check

Deleting Worksheets

Be very cautious when deleting worksheets that contain data. Once a worksheet is deleted, you cannot use the Undo command to bring the sheet back. Deleting a worksheet is a permanent command.

Keyboard Shortcuts

Inserting New Worksheets

- Press the SHIFT key and then the F11 key on your keyboard.

Figure 1.46 shows the final appearance of the GMW Sales workbook.

	A	B	C	D
1	General Merchandise World Retail Sales (in millions)			
2	Month	Unit Sales	Average Price	Sales Dollars
3	January	2,670	$ 9.99	$ 26,685
4	February	2,160	$ 12.49	$ 26,937
5	March	515	$ 14.99	$ 7,701
6	April	590	$ 17.49	$ 10,269
7	May	1,030	$ 14.99	$ 15,405
8	June	2,875	$ 12.49	$ 35,916
9	July	2,700	$ 9.99	$ 26,937
10	August	900	$ 19.99	$ 17,958
11	September	775	$ 19.99	$ 15,708
12	October	1,180	$ 19.99	$ 23,562
13	November	1,800	$ 17.49	$ 31,416
14	December	3,560	$ 14.99	$ 53,370
15	**Total Sales**	**20,755**		**$ 291,864**
16				

Figure 1.46 Final Appearance of the GMW Sales Workbook

Skill Refresher

Renaming Worksheets

1. Double click the worksheet tab.
2. Type the new name.
3. Press the ENTER key.

Moving Worksheets

1. Left click the worksheet tab.
2. Drag it to the desired position.

Deleting Worksheets

1. Open the worksheet to be deleted.
2. Click the Home tab of the Ribbon.
3. Click the down arrow on the Delete button.
4. Select the Delete Sheet option.
5. Click Delete on the warning box.

Key Takeaways

- Formatting skills are critical for creating worksheets that are easy to read and have a professional appearance.
- A series of pound signs (####) in a cell location indicates that the column is too narrow to display the number entered.
- Using the Wrap Text command allows you to stack multiword column headings vertically in a cell location, reducing the need to expand column widths.
- Use the Merge & Center command to center the title of a worksheet directly over the columns that contain data.
- Adding borders or lines will make your worksheet easier to read and helps to separate the data in each column and row.
- You cannot use the Undo command to bring back a worksheet that has been deleted.

Attribution

Adapted by Barbara Lave from How to Use Microsoft Excel:

Media Attributions

- Formatting
- image3-1
- 1-3-fig-1-34
- figure_1-35_formatting_techniques_appllied
- figure_1-38_sheet1_with_data_alignment_features_added
- figure-1-39-title-and-subtitle-added
- 1-40
- 1-3-fig-1-42

STATISTICAL FUNCTIONS

Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker

Learning Objectives

1. Use the SUM function to calculate totals.
2. Use absolute references to calculate percent of totals.
3. Use the COUNT function to count cell locations with numerical values.
4. Use the AVERAGE function to calculate the arithmetic mean.
5. Use the MAX and MIN functions to find the highest and lowest values in a range of cells.

In addition to formulas, another way to conduct mathematical computations in Excel is through functions. Statistical functions apply a mathematical process to a group of cells in a worksheet. For example, the SUM function is used to add the values contained in a range of cells. A list of commonly used statistical functions is shown in **Table 2.4**. Functions are more efficient than formulas when you are applying a mathematical process to a group of cells. If you use a formula to add the values in a range of cells, you would have to add each cell location to the formula one at a time. This can be very time-consuming if you have to add the values in a few hundred cell locations. However, when you use a function, you can highlight all the cells that contain values you wish to sum in just one step. This section demonstrates a variety of statistical functions that we will add to the Personal Budget workbook. In addition to demonstrating functions, this section also reviews percent of total calculations and the use of absolute references.

Table 2.4 Commonly Used Statistical Functions

Function	Output
ABS	The absolute value of a number
AVERAGE	The average or arithmetic mean for a group of numbers
COUNT	The number of cell locations in a range that contain a numeric character
COUNTA	The number of cell locations in a range that contain a text or numeric character
MAX	The highest numeric value in a group of numbers
MEDIAN	The middle number in a group of numbers (half the numbers in the group are higher than the median and half the numbers in the group are lower than the median)
MIN	The lowest numeric value in a group of numbers
MODE	The number that appears most frequently in a group of numbers
PRODUCT	The result of multiplying all the values in a range of cell locations
SQRT	The positive square root of a number
STDEV.S	The standard deviation for a group of numbers based on a sample
SUM	The total of all numeric values in a group

The SUM Function

The SUM function is used when you need to calculate totals for a range of cells or a group of selected cells on a worksheet.

With regard to the **Budget Detail** worksheet, we will use the SUM function to calculate the totals in row 12. It is important to note that there are several methods for adding a function to a worksheet, which will be demonstrated throughout the remainder of this chapter. The following illustrates how a function can be added to a worksheet by typing it into a cell location:

1. Click the **Budget Detail** worksheet tab to open the worksheet.
2. Click cell C12.
3. Type an equal sign =.
4. Type the function name **SUM**.
5. Type an open parenthesis **(**.
6. Click cell C3 and drag down to cell C11. This places the range C3:C11 into the function.
7. Type a closing parenthesis **)**.
8. Press the ENTER key. The function calculates the total for the Monthly Spend column, which is $1,496.

Figure 2.41 shows the appearance of the SUM function added to the **Budget Detail** worksheet before pressing the ENTER key.

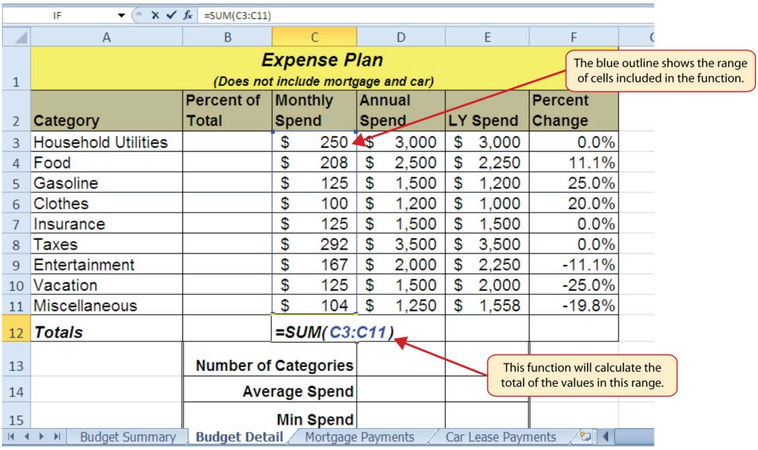


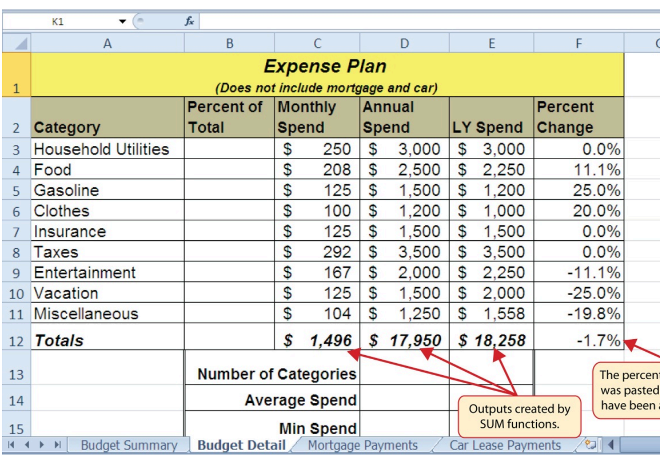
Figure 2.41 Adding the SUM Function to the Budget Detail Worksheet

As shown in **Figure 2.11**, the SUM function was added to cell C12. However, this function is also needed to calculate the totals in the Annual Spend and LY Spend columns. The function can be copied and pasted into these cell locations because of relative referencing. Relative referencing serves the same purpose for functions as it does for formulas. The following demonstrates how the total row is completed:

1. Click cell C12 in the **Budget Detail** worksheet.
2. Click the Copy button in the Home tab of the Ribbon.
3. Highlight cells D12 and E12.
4. Click the Paste button in the Home tab of the Ribbon. This pastes the SUM function into cells D12 and E12 and calculates the totals for these columns.

5. Click cell F11.
6. Click the Copy button in the Home tab of the Ribbon.
7. Click cell F12, then click the Paste button in the Home tab of the Ribbon. Since we now have totals in row 12, we can paste the percent change formula into this row.

Figure 2.42 shows the output of the SUM function that was added to cells C12, D12, and E12. In addition, the percent change formula was copied and pasted into cell F12. Notice that this version of the budget is planning a 1.7% decrease in spending compared to last year.



	A	B	C	D	E	F
1	*Expense Plan* (Does not include mortgage and car)					
2	Category	Percent of Total	Monthly Spend	Annual Spend	LY Spend	Percent Change
3	Household Utilities		$ 250	$ 3,000	$ 3,000	0.0%
4	Food		$ 208	$ 2,500	$ 2,250	11.1%
5	Gasoline		$ 125	$ 1,500	$ 1,200	25.0%
6	Clothes		$ 100	$ 1,200	$ 1,000	20.0%
7	Insurance		$ 125	$ 1,500	$ 1,500	0.0%
8	Taxes		$ 292	$ 3,500	$ 3,500	0.0%
9	Entertainment		$ 167	$ 2,000	$ 2,250	-11.1%
10	Vacation		$ 125	$ 1,500	$ 2,000	-25.0%
11	Miscellaneous		$ 104	$ 1,250	$ 1,558	-19.8%
12	*Totals*		*$ 1,496*	*$ 17,950*	*$ 18,258*	-1.7%
13		Number of Categories				
14		Average Spend				
15		Min Spend				

Figure 2.42 Results of the SUM Function in the Budget Detail Worksheet

Integrity Check

Cell Ranges in Statistical Functions

When you intend to use a statistical function on a range of cells in a worksheet, make sure there are two cell locations separated by a colon and not a comma. If you enter two cell locations separated by a comma, the function will produce an output but it will be applied to only two cell locations instead of a range of cells. For example, the SUM function shown in **Figure 2.13** will add only the values in cells C3 and C11, not the range C3:C11.

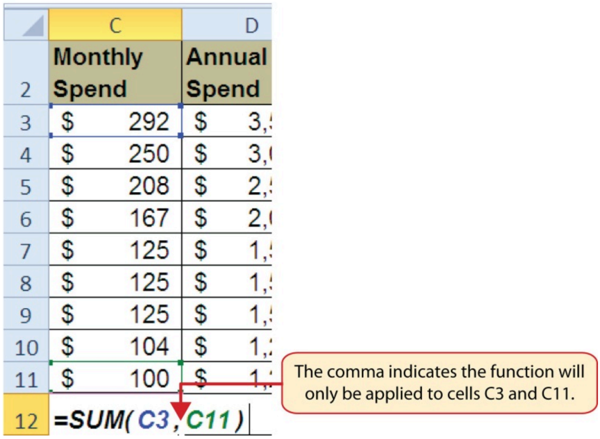


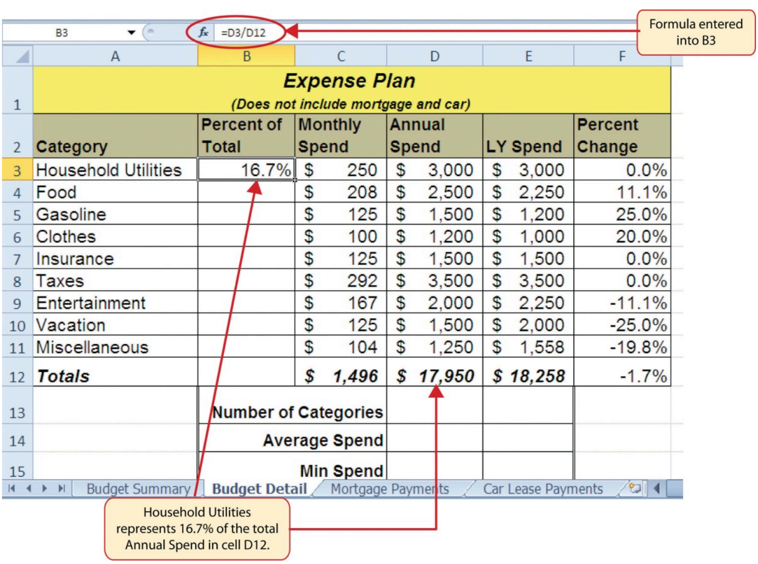
Figure 2.42 SUM Function Adding Two Cell Locations

Absolute References (Calculating Percent of Totals)

Data file: Continue with CH2 Personal Budget.

Since totals were added to row 12 of the **Budget Detail** worksheet, a percent of total calculation can be added to Column B beginning in cell B3. The percent of total calculation shows the percentage for each value in the Annual Spend column with respect to the total in cell D12. However, after the formula is created, it will be necessary to turn off Excel's relative referencing feature before copying and pasting the formula to the rest of the cell locations in the column. Turning off Excel's relative referencing feature is accomplished through an absolute reference. The following steps explain how this is done:

1. Click cell B3 in the **Budget Detail** worksheet.
2. Type an equal sign =.
3. Click cell D3.
4. Type a forward slash /.
5. Click cell D12.
6. Press the ENTER key. You will see that Household Utilities represents 16.7% of the Annual Spend budget (see **Figure 2.44**).



	A	B	C	D	E	F
1	Expense Plan (Does not include mortgage and car)					
2	Category	Percent of Total	Monthly Spend	Annual Spend	LY Spend	Percent Change
3	Household Utilities	16.7%	$ 250	$ 3,000	$ 3,000	0.0%
4	Food		$ 208	$ 2,500	$ 2,250	11.1%
5	Gasoline		$ 125	$ 1,500	$ 1,200	25.0%
6	Clothes		$ 100	$ 1,200	$ 1,000	20.0%
7	Insurance		$ 125	$ 1,500	$ 1,500	0.0%
8	Taxes		$ 292	$ 3,500	$ 3,500	0.0%
9	Entertainment		$ 167	$ 2,000	$ 2,250	-11.1%
10	Vacation		$ 125	$ 1,500	$ 2,000	-25.0%
11	Miscellaneous		$ 104	$ 1,250	$ 1,558	-19.8%
12	Totals		$ 1,496	$ 17,950	$ 18,258	-1.7%
13		Number of Categories				
14		Average Spend				
15		Min Spend				

Figure 2.44 Adding a Formula to Calculate the Percent of Total

Figure 2.44 shows the completed formula that is calculating the percentage that Household Utilities Annual Spend represents to the total Annual Spend for the budget (see cell B3). Normally, we would copy this formula and paste it into the range B4:B11. However, because of relative referencing, both cell references will increase by one row as the formula is pasted into the cells below B3. This is fine for the first cell reference in the formula (D3) but not for the second cell reference (D12). **Figure 2.45** illustrates what happens if we paste the formula into the range B4:B12 in its current state. Notice that Excel produces the #DIV/0 error code. This means that Excel is trying to divide a number by zero, which is impossible. Looking at the formula in cell B4, you see that the first cell reference was changed from D3 to D4. This is fine because we now want to divide the Annual Spend for Insurance by the total Annual Spend in cell D12. However, Excel has also changed the D12 cell reference to D13. Because cell location D13 is blank, the formula produces the #DIV/0 error code.

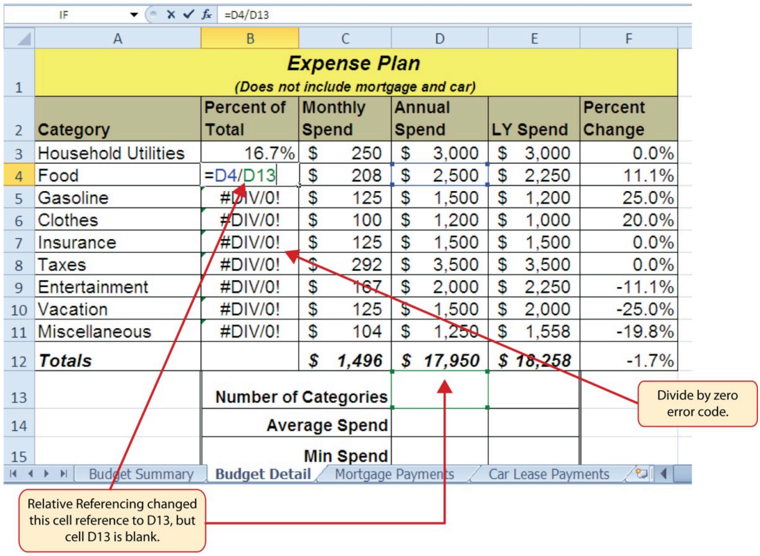


Figure 2.45 #DIV/0 Error from Relative Referencing

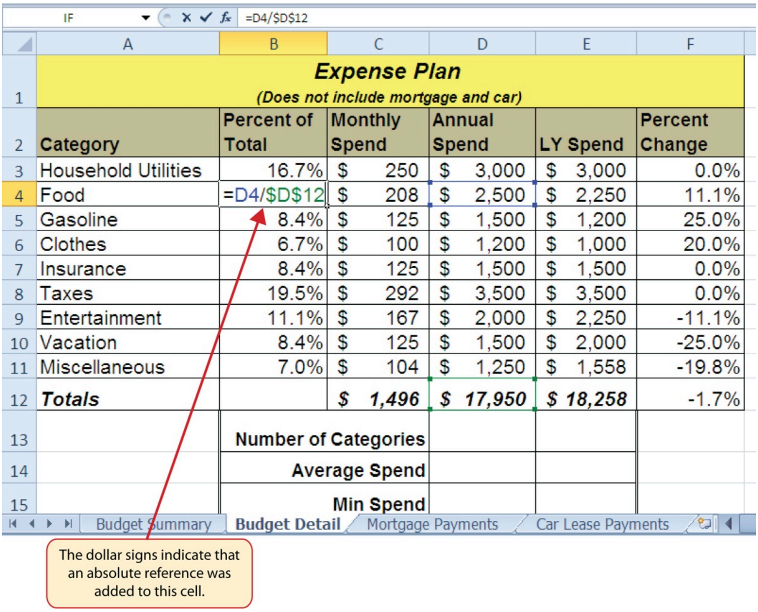
To eliminate the divide-by-zero error shown in **Figure 2.15** we must add an absolute reference to cell D12 in the formula. An absolute reference prevents relative referencing from changing a cell reference in a formula. This is also referred to as locking a cell. The following explains how this is accomplished:

1. Double click cell B3.
2. Place the mouse pointer in front of D12 and click. The blinking cursor should be in front of the D in the cell reference D12.
3. Press the F4 key. You will see a dollar sign ($) added in front of the column letter D and the row number 12.

You can also type the dollar signs in front of the column letter and row number.

4. Press the ENTER key.
5. Click cell B3.
6. Click the Copy button in the Home tab of the Ribbon.
7. Highlight the range B4:B11.
8. Click the Paste button in the Home tab of the Ribbon.

Figure 2.46 shows the percent of total formula with an absolute reference added to D12. Notice that in cell B4, the cell reference remains D12 instead of changing to D13 as shown in **Figure 2.15**. Also, you will see that the percentages are being calculated in the rest of the cells in the column, and the divide-by-zero error is now eliminated.



	A	B	C	D	E	F
1	**Expense Plan** (Does not include mortgage and car)					
2	Category	Percent of Total	Monthly Spend	Annual Spend	LY Spend	Percent Change
3	Household Utilities	16.7%	$ 250	$ 3,000	$ 3,000	0.0%
4	Food	=D4/D12	$ 208	$ 2,500	$ 2,250	11.1%
5	Gasoline	8.4%	$ 125	$ 1,500	$ 1,200	25.0%
6	Clothes	6.7%	$ 100	$ 1,200	$ 1,000	20.0%
7	Insurance	8.4%	$ 125	$ 1,500	$ 1,500	0.0%
8	Taxes	19.5%	$ 292	$ 3,500	$ 3,500	0.0%
9	Entertainment	11.1%	$ 167	$ 2,000	$ 2,250	-11.1%
10	Vacation	8.4%	$ 125	$ 1,500	$ 2,000	-25.0%
11	Miscellaneous	7.0%	$ 104	$ 1,250	$ 1,558	-19.8%
12	*Totals*		*$ 1,496*	*$ 17,950*	*$ 18,258*	-1.7%
13			Number of Categories			
14			Average Spend			
15			Min Spend			

Figure 2.46 Adding an Absolute Reference to a Cell Reference in a Formula

Skill Refresher

Absolute References

1. Click in front of the column letter of a cell reference in a formula or function that you do not want altered when the formula or function is pasted into a new cell location.

2. Press the F4 key or type a dollar sign $ in front of the column letter and row number of the cell reference.

The COUNT Function

Data file: Continue with CH2 Personal Budget.

The next function that we will add to the **Budget Detail** worksheet is the COUNT function. The COUNT function is used to determine how many cells in a range contain a numeric entry. The COUNT function will not work for counting text or other non-numeric entries. For the **Budget Detail** worksheet, we will use the COUNT function to count the number of items that are planned in the Annual Spend column (Column D). The following explains how the COUNT function is added to the worksheet by using the function list:

1. Click cell D13 in the Budget Detail worksheet.
2. Type an equal sign =.
3. Type the letter **C**.
4. Click the down arrow on the scroll bar of the function list (see **Figure 2.47**) and find the word COUNT.
5. Double click the word COUNT from the function list.

6. Highlight the range D3:D11.
7. You can type a closing parenthesis **)** and then press the ENTER key, or simply press the ENTER key and Excel will close the function for you. The function produces an output of 9 since there are 9 items planned on the worksheet.

Figure 2.47 shows the function list box that appears after completing steps 2 and 3 for the COUNT function. The function list provides an alternative method for adding a function to a worksheet.

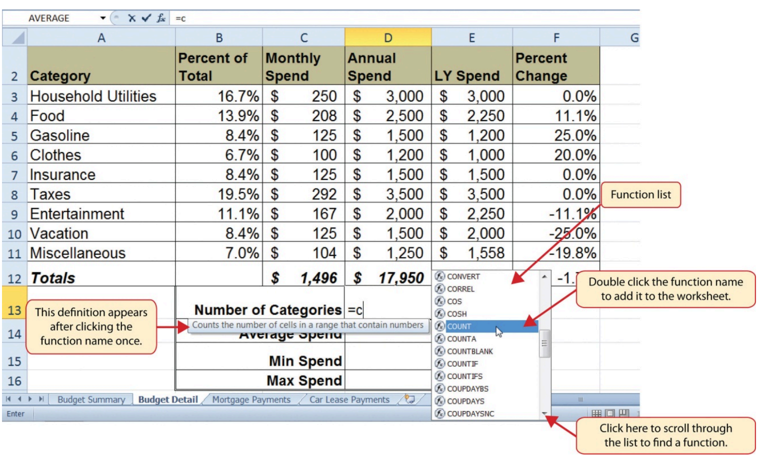
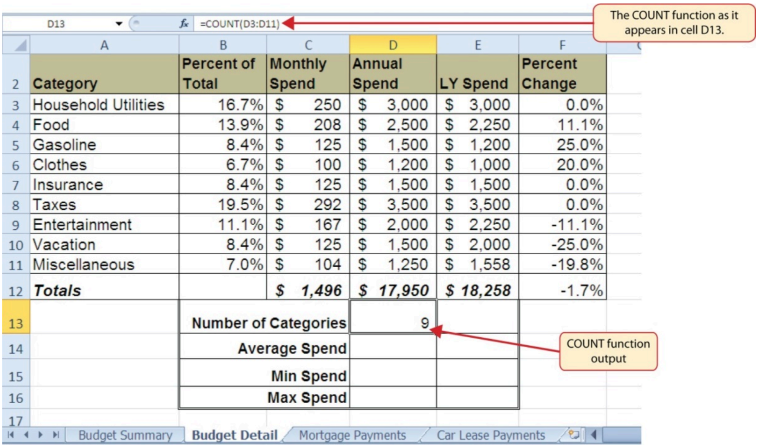


Figure 2.47 Using the Function List to Add the COUNT Function

Figure 2.48 shows the output of the COUNT function after pressing the ENTER key. The function counts the number of cells in the range D3:D11 that contain a numeric value. The result of 9 indicates that there are 9 categories planned for this budget.



	A	B	C	D	E	F
2	Category	Percent of Total	Monthly Spend	Annual Spend	LY Spend	Percent Change
3	Household Utilities	16.7%	$ 250	$ 3,000	$ 3,000	0.0%
4	Food	13.9%	$ 208	$ 2,500	$ 2,250	11.1%
5	Gasoline	8.4%	$ 125	$ 1,500	$ 1,200	25.0%
6	Clothes	6.7%	$ 100	$ 1,200	$ 1,000	20.0%
7	Insurance	8.4%	$ 125	$ 1,500	$ 1,500	0.0%
8	Taxes	19.5%	$ 292	$ 3,500	$ 3,500	0.0%
9	Entertainment	11.1%	$ 167	$ 2,000	$ 2,250	-11.1%
10	Vacation	8.4%	$ 125	$ 1,500	$ 2,000	-25.0%
11	Miscellaneous	7.0%	$ 104	$ 1,250	$ 1,558	-19.8%
12	*Totals*		*$ 1,496*	*$ 17,950*	*$ 18,258*	-1.7%
13		Number of Categories		9		
14		Average Spend				
15		Min Spend				
16		Max Spend				

Figure 2.48 Completed COUNT Function in the Budget Detail Worksheet

The AVERAGE Function

The next function we will add to the **Budget Detail** worksheet is the AVERAGE function. This function is used to calculate the ***arithmetic mean*** for a group of numbers. For the **Budget Detail** worksheet, we will use the function to calculate the average of the values in the Annual Spend column. We will add this to the worksheet by using the Function Library. The following steps explain how this is accomplished:

1. Click cell D14 in the **Budget Detail** worksheet.

2. Click the Formulas tab on the Ribbon.
3. Click the More Functions button in the Function Library group of commands.
4. Place the mouse pointer over the Statistical option from the drop-down list of options.
5. Click the AVERAGE function name from the list of functions that appear in the menu (see **Figure 2.49**). This opens the Function Arguments dialog box.
6. Click the Collapse Dialog button in the Function Arguments dialog box (see **Figure 2.410**).
7. Highlight the range D3:D11.
8. Click the Expand Dialog button in the Function Arguments dialog box (see **Figure 2.411**). You can also press the ENTER key to get the same result.
9. Click the OK button on the Function Arguments dialog box. This adds the AVERAGE function to the worksheet.

Figure 2.49 illustrates how a function is selected from the Function Library in the Formulas tab of the Ribbon.

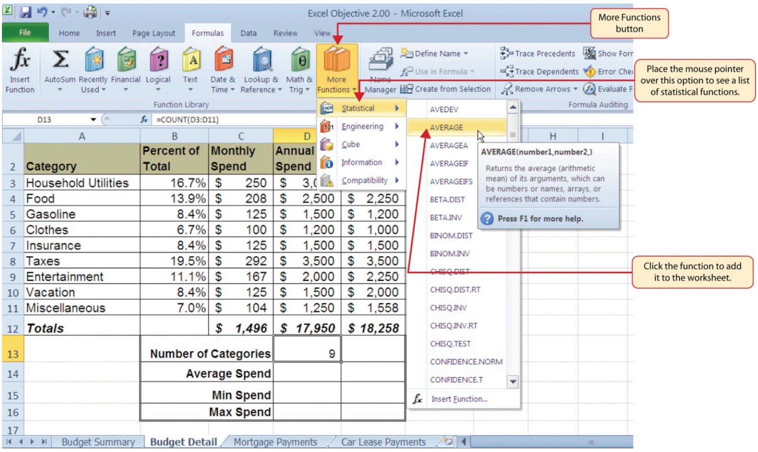


Figure 2.19 Selecting the AVERAGE Function from the Function Library

Figure 2.410 shows the Function Arguments dialog box. This appears after a function is selected from the Function Library. The Collapse Dialog button is used to hide the dialog box so a range of cells can be highlighted on the worksheet and then added to the function.

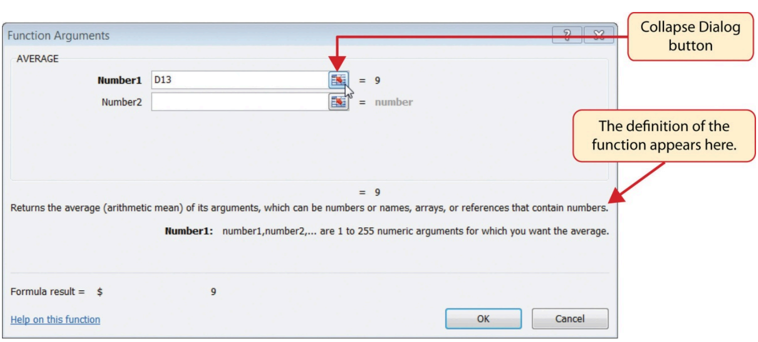


Figure 2.410 Function Arguments Dialog Box

Figure 2.411 shows how a range of cells can be selected from the Function Arguments dialog box once it has been collapsed.

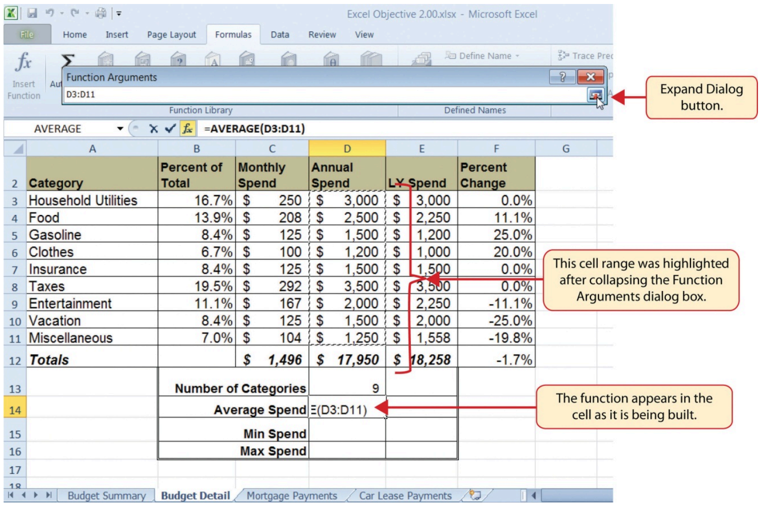


Figure 2.411 Selecting a Range from the Function Arguments Dialog Box

Figure 2.412 shows the Function Arguments dialog box after the cell range is defined for the AVERAGE function. The dialog box shows the result of the function before it is added to the cell location. This allows you to assess the function output to determine whether it makes sense before adding it to the worksheet.

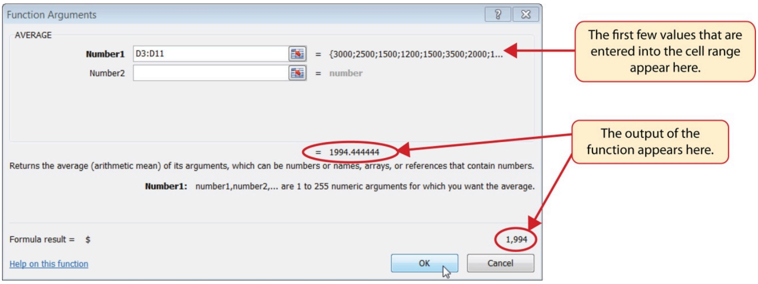
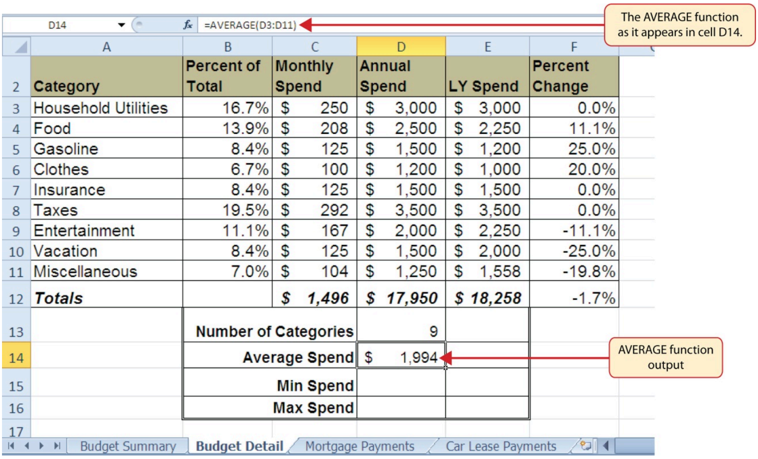


Figure 2.412 Function Arguments Dialog Box after a Cell Range Is Defined for a Function

Figure 2.413 shows the completed AVERAGE function in the **Budget Detail** worksheet. The output of the function shows that on average we expect to spend $1,994 for each of the categories listed in Column A of the budget. This average spend calculation per category can be used as an indicator to determine which categories are costing more or less than the average budgeted spend dollars.



D14 =AVERAGE(D3:D11)

	A	B	C	D	E	F
2	Category	Percent of Total	Monthly Spend	Annual Spend	LY Spend	Percent Change
3	Household Utilities	16.7%	$ 250	$ 3,000	$ 3,000	0.0%
4	Food	13.9%	$ 208	$ 2,500	$ 2,250	11.1%
5	Gasoline	8.4%	$ 125	$ 1,500	$ 1,200	25.0%
6	Clothes	6.7%	$ 100	$ 1,200	$ 1,000	20.0%
7	Insurance	8.4%	$ 125	$ 1,500	$ 1,500	0.0%
8	Taxes	19.5%	$ 292	$ 3,500	$ 3,500	0.0%
9	Entertainment	11.1%	$ 167	$ 2,000	$ 2,250	-11.1%
10	Vacation	8.4%	$ 125	$ 1,500	$ 2,000	-25.0%
11	Miscellaneous	7.0%	$ 104	$ 1,250	$ 1,558	-19.8%
12	*Totals*		*$ 1,496*	*$ 17,950*	*$ 18,258*	-1.7%
13		Number of Categories		9		
14		Average Spend		$ 1,994		
15		Min Spend				
16		Max Spend				

Figure 2.413 Completed AVERAGE Function

The MAX and MIN Functions

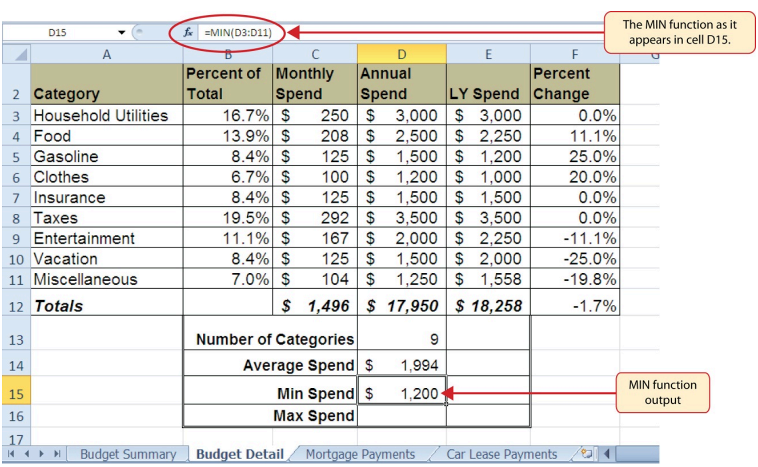
Data file: Continue with CH2 Personal Budget.

The final two statistical functions that we will add to the **Budget Detail** worksheet are the **MAX** and **MIN** functions. These functions identify the highest and lowest values in a range of cells. The following steps explain how to add these functions to the **Budget Detail** worksheet:

1. Click cell D15 in the **Budget Det**ail worksheet.
2. Type an equal sign =.
3. Type the word **MIN**.
4. Type an open parenthesis **(**.
5. Highlight the range D3:D11.
6. Type a closing parenthesis **)** and press the ENTER key,

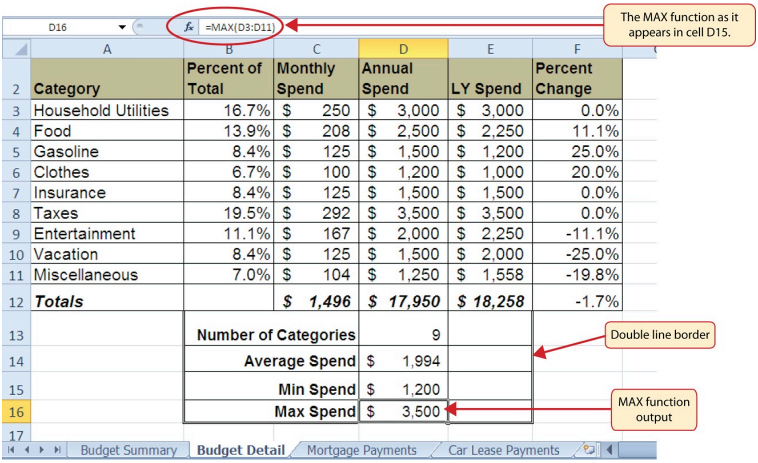
or simply press the ENTER key and Excel will close the function for you. The MIN function produces an output of $1,200, which is the lowest value in the Annual Spend column (see **Figure 2.414**).

7. Click cell D16.
8. Type an equal sign =.
9. Type the word **MAX**.
10. Type an open parenthesis **(**.
11. Highlight the range D3:D11.
12. Type a closing parenthesis **)** and press the ENTER key, or simply press the ENTER key and Excel will close the function for you. The MAX function produces an output of $3,500. This is the highest value in the Annual Spend column (see **Figure 2.415**).



	A	B	C	D	E	F
2	Category	Percent of Total	Monthly Spend	Annual Spend	LY Spend	Percent Change
3	Household Utilities	16.7%	$ 250	$ 3,000	$ 3,000	0.0%
4	Food	13.9%	$ 208	$ 2,500	$ 2,250	11.1%
5	Gasoline	8.4%	$ 125	$ 1,500	$ 1,200	25.0%
6	Clothes	6.7%	$ 100	$ 1,200	$ 1,000	20.0%
7	Insurance	8.4%	$ 125	$ 1,500	$ 1,500	0.0%
8	Taxes	19.5%	$ 292	$ 3,500	$ 3,500	0.0%
9	Entertainment	11.1%	$ 167	$ 2,000	$ 2,250	-11.1%
10	Vacation	8.4%	$ 125	$ 1,500	$ 2,000	-25.0%
11	Miscellaneous	7.0%	$ 104	$ 1,250	$ 1,558	-19.8%
12	*Totals*		*$ 1,496*	*$ 17,950*	*$ 18,258*	-1.7%
13		Number of Categories		9		
14		Average Spend		$ 1,994		
15		Min Spend		$ 1,200		
16		Max Spend				
17						

Figure 2.414 MIN Function Added to the Budget Detail Worksheet



Category	Percent of Total	Monthly Spend	Annual Spend	LY Spend	Percent Change
Household Utilities	16.7%	$ 250	$ 3,000	$ 3,000	0.0%
Food	13.9%	$ 208	$ 2,500	$ 2,250	11.1%
Gasoline	8.4%	$ 125	$ 1,500	$ 1,200	25.0%
Clothes	6.7%	$ 100	$ 1,200	$ 1,000	20.0%
Insurance	8.4%	$ 125	$ 1,500	$ 1,500	0.0%
Taxes	19.5%	$ 292	$ 3,500	$ 3,500	0.0%
Entertainment	11.1%	$ 167	$ 2,000	$ 2,250	-11.1%
Vacation	8.4%	$ 125	$ 1,500	$ 2,000	-25.0%
Miscellaneous	7.0%	$ 104	$ 1,250	$ 1,558	-19.8%
Totals		***$ 1,496***	***$ 17,950***	***$ 18,258***	-1.7%
	Number of Categories		9		
	Average Spend		$ 1,994		
	Min Spend		$ 1,200		
	Max Spend		$ 3,500		

Figure 2.415 MAX Function Added to the Budget Detail Worksheet

Skill Refresher

Statistical Functions

1. Type an equal sign **=**.
2. Type the function name followed by an open parenthesis **(** or double click the function name from the function list.
3. Highlight a range on a worksheet or click individual cell locations followed by commas.
4. Type a closing parenthesis **)** and press the ENTER key or press the ENTER key to close

the function.

Key Takeaways

- Statistical functions are used when a mathematical process is required for a range of cells, such as summing the values in several cell locations. For these computations, functions are preferable to formulas because adding many cell locations one at a time to a formula can be very time-consuming.
- Statistical functions can be created using cell ranges or selected cell locations separated by commas. Make sure you use a cell range (two cell locations separated by a colon) when applying a statistical function to a contiguous range of cells.
- To prevent Excel from changing the cell references in a formula or function when they are pasted to a new cell location, you must use an absolute reference. You can do this by placing a dollar sign ($) in front of the column

letter and row number of a cell reference.

- The #DIV/0 error appears if you create a formula that attempts to divide a constant or the value in a cell reference by zero.

Attribution

FORMULAS

Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker

Learning Objectives

- Learn how to create basic formulas.
- Understand relative referencing when copying and pasting formulas.
- Work with complex formulas by controlling the order of mathematical operations.

This section reviews the fundamental skills for entering formulas into an Excel worksheet. The objective used for this chapter is the construction of a personal cash budget. Most financial advisors recommend that all households construct and maintain a personal budget to achieve and maintain strong financial health. Organizing and maintaining a personal budget is a skill you can practice at any point in your life. Whether you are managing your expenses during college or maintaining the finances of a family of four, a personal budget can be a vital tool when making financial decisions. Excel can make managing your money a fun and rewarding exercise.

Figure 2.51 shows the completed workbook that will be demonstrated in this chapter. Notice that this workbook contains four worksheets. The first worksheet, **Budget Summary**, contains formulas that utilize or reference the data in the other three worksheets. As a result, the **Budget Summary** worksheet serves as an overview of the data that was entered and calculated in the other three worksheets of the workbook.

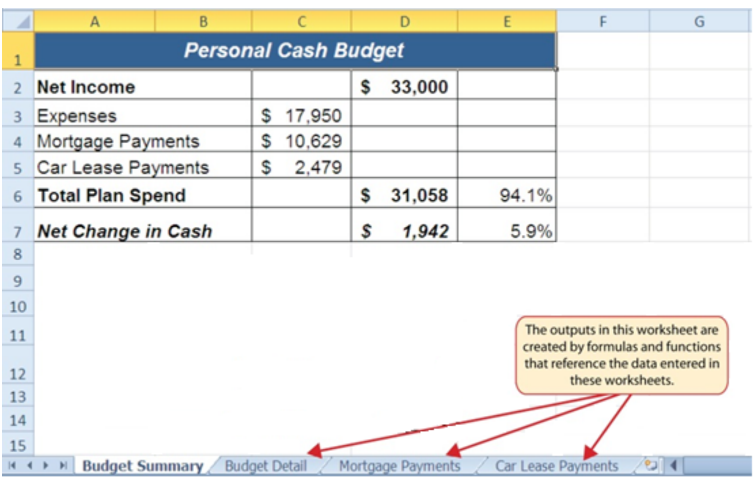


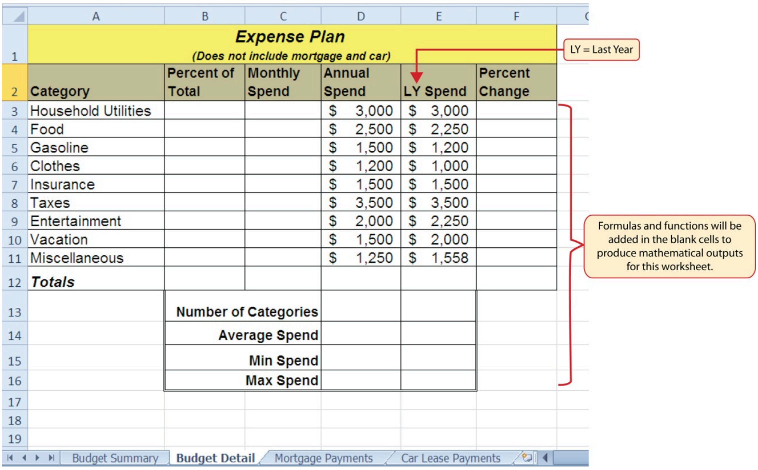
Figure 2.51 Completed Personal Budget Workbook

Creating a Basic Formula

If you wish to follow along, download Data File: [CH2 Data]

Formulas are used to calculate a variety of mathematical outputs in Excel and can be used to create virtually any custom calculation required for your objective. Furthermore, when constructing a formula in Excel, you use cell locations that, when added to a formula, become cell references. This means that Excel uses, or references, the number entered into the cell location when calculating a mathematical output. As a result, when the numbers in the cell references are changed, Excel automatically produces a new output. This is what gives Excel the ability to create a variety of what-if scenarios, which will be explained later in the chapter.

To demonstrate the construction of a basic formula, we will begin working on the **Budget Detail** worksheet in the Personal Budget workbook, which is shown in **Figure 2.52**. To complete this worksheet, we will add several formulas and functions. **Table 2.51** provides definitions for each of the spend categories listed in the range A3:A11. When you develop a personal budget, these categories are defined on the basis of how you spend your money. It is likely that every person could have different categories or define the same categories differently. Therefore, it is important to review the definitions in **Table 2.51** to understand how we are defining these categories before proceeding.



	A	B	C	D	E	F
1	***Expense Plan*** *(Does not include mortgage and car)*					
2	**Category**	**Percent of Total**	**Monthly Spend**	**Annual Spend**	**LY Spend**	**Percent Change**
3	Household Utilities			$ 3,000	$ 3,000	
4	Food			$ 2,500	$ 2,250	
5	Gasoline			$ 1,500	$ 1,200	
6	Clothes			$ 1,200	$ 1,000	
7	Insurance			$ 1,500	$ 1,500	
8	Taxes			$ 3,500	$ 3,500	
9	Entertainment			$ 2,000	$ 2,250	
10	Vacation			$ 1,500	$ 2,000	
11	Miscellaneous			$ 1,250	$ 1,558	
12	***Totals***					
13		**Number of Categories**				
14		**Average Spend**				
15		**Min Spend**				
16		**Max Spend**				
17						
18						
19						

Figure 2.52 Budget Detail Worksheet

Table 2.51 Spend Category Definitions

Category	Definition
Household Utilities	Money spent on electricity, heat, and water and on cable, phone, and Internet access
Food	Money spent on groceries, toiletries, and related items
Gasoline	Money spent on fuel for automobiles
Clothes	Money spent on clothes, shoes, and accessories
Insurance	Money spent on homeowner's or automobile insurance
Taxes	Money spent on school and property taxes (this example of the personal budget assumes that we own property).
Entertainment	Money spent on entertainment, including dining out, movie and theater tickets, parties, and so on
Vacation	Money spent on vacations
Miscellaneous	Includes any other spending categories, such as textbooks, software, journals, school or work supplies, and so on

The first formula that we will add to the **Budget Detail** worksheet will calculate the Monthly Spend values. The formula will be constructed so that it takes the values in the Annual Spend column and divides them by 12. This will show how much money will be spent per month for each of the categories listed in Column A. The following explains how this formula is created:

1. Open the Data file named **CH2 Data** and use the File/ Save As command to save it with the new name **CH2 Personal Budget**.
2. Click the **Budget Detail** worksheet tab to open the worksheet.
3. Click cell C3.
4. Type an equal sign =. When the first character entered into a cell location is an equal sign, it signals Excel to perform a calculation or produce a logical output.
5. Type **D3**. This adds D3 to the formula, which is now a cell reference. Excel will use whatever value is entered into cell D3 to produce an output.
6. Type the slash symbol /. This is the symbol for division in Excel. As shown in **Table 2.52** the mathematical operators in Excel are slightly different from those found on a typical calculator.
7. Type the number **12**. This divides the value in cell D3 by 12. In this formula, a number, or constant, is used instead of a cell reference because it will not change. In other words, there will always be 12 months in a year.
8. Press the ENTER key.

Table 2.2 Excel Mathematical Operators

Symbol	Operation
+	Addition
−	Subtraction
/	Division
*	Multiplication
^	Power/Exponent

Why?

Use Cell References

Cell references enable Excel to dynamically produce new outputs when one or more inputs in the referenced cells are changed. Cell references also allow you to trace how outputs are being calculated in a formula. As a result, you should never use a calculator to determine a mathematical output and type it into the cell location of a worksheet. Doing so eliminates Excel's cell-referencing benefits as well as your ability to trace a formula to determine how outputs are being produced.

Figure 2.35 shows how the formula appears in cell C3 before you press the ENTER key. **Figure 2.54** shows the output of the formula after you press the ENTER key. The Monthly Spend for Household Utilities is $250 because the formula is taking the Annual Spend in cell D3 and dividing it by 12. If the value in cell D3 is changed, the formula automatically produces a new output. We are calculating the spend per month for each category because people often get paid and are billed for these items on a monthly basis. This formula allows you to compare your monthly income to your monthly bills to determine whether you have enough income to pay these expenses.

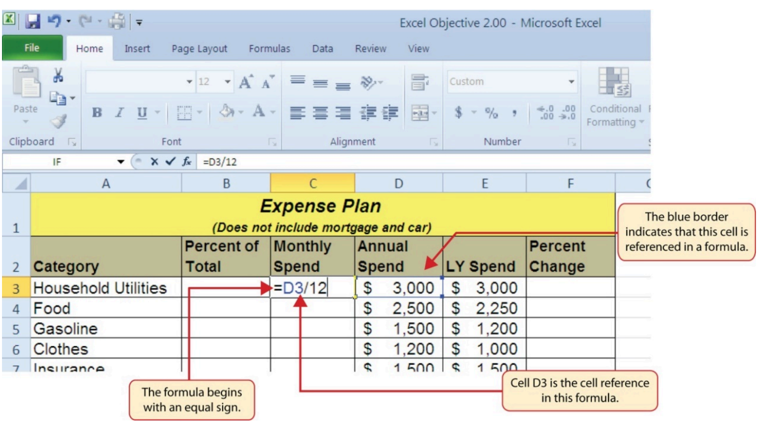


Figure 2.53 Adding a Formula to a Worksheet

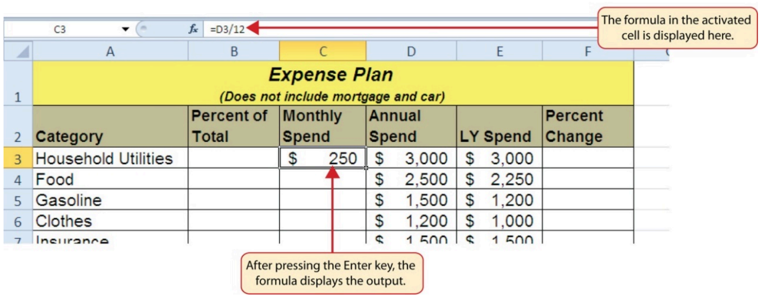


Figure 2.54 Formula Output for Monthly Spend

Why?

Use Universal Constants

If you are using constants, or numerical values, in an Excel formula, they should be universal constants that do not change, such as the number of days in a week, weeks in a year, and so on. Do not type the values that exist in cell locations into an Excel formula. This will eliminate Excel's cell-referencing benefits, which means if the value in the cell location you are using in a formula is changed, Excel will not be able to produce a new output.

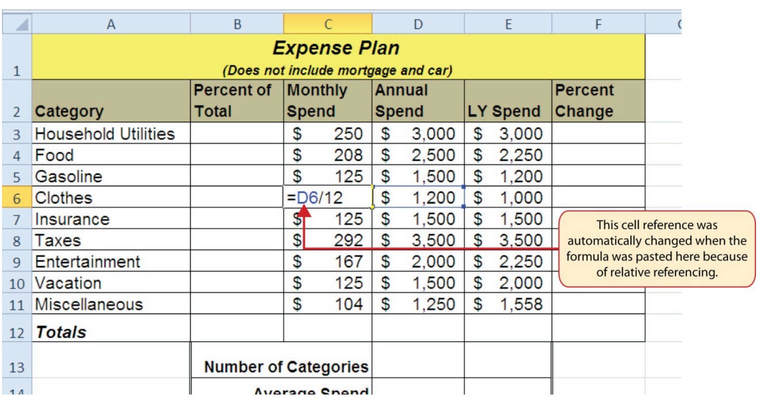
Relative References (Copying and Pasting Formulas)

Once a formula is typed into a worksheet, it can be copied and pasted to other cell locations. For example, **Figure 2.54** shows the output of the formula that was entered into cell C3. However, this calculation needs to be performed for the rest of the cell locations in Column C. Since we used the D3 cell reference in the formula, Excel automatically adjusts that cell reference when the formula is copied and pasted into the rest of the cell locations in the column. This is called relative referencing and is demonstrated as follows:

1. Click cell C3.
2. Place the mouse pointer over the Auto Fill Handle.
3. When the mouse pointer turns from a white block plus sign to a black plus sign, click and drag down to cell C11. This pastes the formula into the range C4:C11.
4. Double click cell C6. Notice that the cell reference in the formula is automatically changed to D6.

5. Press the ENTER key.

Figure 2.55 shows the outputs added to the rest of the cell locations in the Monthly Spend column. For each row, the formula takes the value in the Annual Spend column and divides it by 12. You will also see that cell D6 has been double clicked to show the formula. Notice that Excel automatically changed the original cell reference of D3 to D6. This is the result of relative referencing, which means Excel automatically adjusts a cell reference relative to its original location when it is pasted into new cell locations. In this example, the formula was pasted into eight cell locations below the original cell location. As a result, Excel increased the row number of the original cell reference by a value of one for each row it was pasted into.



	A	B	C	D	E	F
1	*Expense Plan* (Does not include mortgage and car)					
2	Category	Percent of Total	Monthly Spend	Annual Spend	LY Spend	Percent Change
3	Household Utilities		$ 250	$ 3,000	$ 3,000	
4	Food		$ 208	$ 2,500	$ 2,250	
5	Gasoline		$ 125	$ 1,500	$ 1,200	
6	Clothes		=D6/12	$ 1,200	$ 1,000	
7	Insurance		$ 125	$ 1,500	$ 1,500	
8	Taxes		$ 292	$ 3,500	$ 3,500	
9	Entertainment		$ 167	$ 2,000	$ 2,250	
10	Vacation		$ 125	$ 1,500	$ 2,000	
11	Miscellaneous		$ 104	$ 1,250	$ 1,558	
12	*Totals*					
13		Number of Categories				

Figure 2.55 Relative Reference Example

Why?

Use Relative Referencing

Relative referencing is a convenient feature in Excel. When you use cell references in a formula, Excel automatically adjusts the cell references when the formula is pasted into new cell locations. If this feature were not available, you would have to manually retype the formula when you want the same calculation applied to other cell locations in a column or row.

Creating Complex Formulas (Controlling the Order of Operations)

The next formula to be added to the Personal Budget workbook is the percent change over last year. This formula determines the difference between the values in the LY (Last Year) Spend column and shows the difference in terms of a percentage. This requires that the order of mathematical operations be controlled to get an accurate result. **Table 2.53** shows the standard order of operations for a typical formula. To change the order of operations shown in the table, we use parentheses to process certain mathematical calculations first. This formula is added to the worksheet as follows:

1. Click cell F3 in the **Budget Detail** worksheet.
2. Type an equal sign =.
3. Type an open parenthesis **(**.
4. Click cell D3. This will add a cell reference to cell D3 to the formula. When building formulas, you can click cell locations instead of typing them.
5. Type a minus sign −.
6. Click cell E3 to add this cell reference to the formula.
7. Type a closing parenthesis **)**.
8. Type the slash / symbol for division.
9. Click cell E3. This completes the formula that will calculate the percent change of last year's actual spent dollars vs. this year's budgeted spend dollars (**see Figure 2.6**).
10. Press the ENTER key.
11. Click cell F3 to activate it.
12. Place the mouse pointer over the Auto Fill Handle.
13. When the mouse pointer turns from a white block plus sign to a black plus sign, click and drag down to cell F11. This pastes the formula into the range F4:F11.

Table 2.53 Standard Order of Mathematical Operations

Symbol	Order
( )	Override Standard Order: Any mathematical computations placed in parentheses are performed first and override the standard order of operations. If there are layers of parentheses used in a formula, Excel computes the innermost parentheses first and the outermost parentheses last.
^	First: Excel executes any exponential computations first.
* or /	Second: Excel performs any multiplication or division computations second. When there are multiple instances of these computations in a formula, they are executed in order from left to right.
+ or –	Third: Excel performs any addition or subtraction computations third. When there are multiple instances of these computations in a formula, they are executed in order from left to right.

Figure 2.56 shows the formula that was added to the **Budget Detail** worksheet to calculate the percent change in spending. The parentheses were added to this formula to control the order of operations. Any mathematical computations placed in parentheses are executed first before the standard order of mathematical operations (see **Table 2.53)**. In this case, if parentheses were not used, Excel would produce an erroneous result for this worksheet.

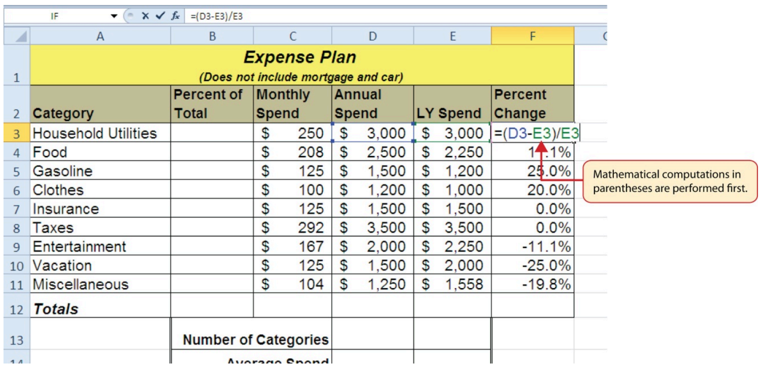


Figure 2.56 Adding the Percent Change Formula

Figure 2.57 shows the result of the percent change formula if the parentheses are removed. The formula produces a result of a 299900% increase. Since there is no change between the LY spend and the budget Annual Spend, the result should be 0%. However, without the parentheses, Excel is following the standard order of operations. This means the value in cell E3 will be divided by E3 first (3,000/3,000), which is 1. Then, the value of 1 will be subtracted from the value in cell D3 (3,000−1), which is 2,999. Since cell F3 is formatted as a percentage, Excel expresses the output as an increase of 299900%.

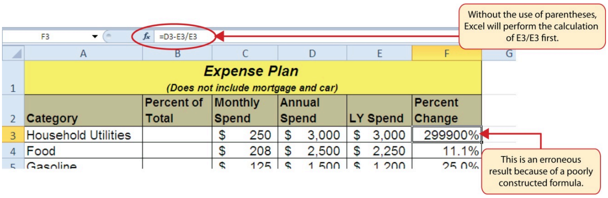


Figure 2.57 Removing the Parentheses from the Percent Change Formula

Integrity Check

Does the Output of Your Formula Make Sense?

It is important to note that the accuracy of the output produced by a formula depends on how it is constructed. Therefore, always check the result of your formula to see whether it makes sense with data in your worksheet. As shown in **Figure 2.57**, a poorly constructed formula can give you an inaccurate result. In other words, you can see that there is no change between the Annual

Spend and LY Spend for Household Utilities. Therefore, the result of the formula should be 0%. However, since the parentheses were removed in this case, the formula is clearly producing an erroneous result.

Skill Refresher

Formulas

1. Type an equal sign **=**.
2. Click or type a cell location. If using constants, type a number.
3. Type a mathematical operator.
4. Click or type a cell location. If using constants, type a number.
5. Use parentheses where necessary to control the order of operations.
6. Press the ENTER key

Key Takeaways

- Mathematical computations are conducted through formulas and functions.
- An equal sign **=** precedes all formulas and functions.
- Formulas and functions must be created with cell references to conduct what-if scenarios where mathematical outputs are recalculated when one or more inputs are changed.
- Mathematical operators on a typical calculator are different from those used in Excel. **Table 2.2** "Excel Mathematical Operators" lists Excel mathematical operators.
- When using numerical values in formulas and functions, only use universal constants that do not change, such as days in a week, months in a year, and so on.
- Relative referencing automatically adjusts the cell references in formulas and functions when they are pasted into new locations on a worksheet. This eliminates the need to retype formulas and functions when they are needed in multiple rows or columns on a worksheet.
- Parentheses must be used to control the order

of operations when necessary for complex formulas.

Attribution

CHOOSING A CHART TYPE

Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, Diane Shingledecker, and Gail Poitrast

LEARNING OBJECTIVES

1. Construct a line chart to show a time series trend.
2. Learn how to adjust the Y axis scale.
3. Learn how to use a column chart to show a frequency distribution.
4. Learn how to use a pie chart to show the percent of total for a data set.
5. Construct a stacked column chart to show how a percent of total changes over time.
6. Construct a histogram when given quantitative data.

This section reviews the most commonly used Excel chart types. To demonstrate the variety of chart types available in Excel, it is necessary to use a variety of data sets. This is necessary not only to demonstrate the construction of charts but also to explain how to choose the right type of chart given your data and the idea you intend to communicate.

Choosing a Chart Type

Before we begin, let's review a few key points you need to consider before creating any chart in Excel.

1. **The first is identifying your idea or message**. It is important to keep in mind that the primary purpose of a chart is to present quantitative information to an audience. Therefore, you must first decide what message or idea you wish to present. This is critical in helping you select specific data from a worksheet that will be used in a chart. Throughout this chapter, we will reinforce the intended message first before creating each chart.
2. **The second key point is selecting the right chart type**. The chart type you select will depend on the data you have and the message you intend to communicate. If you are using categorical (sometimes referred to as qualitative data), a bar chart or pie chart is appropriate. If you are using quantitative data, then a histogram or line chart is appropriate.

3. **The third key point is identifying the values that should appear on the X and Y axes**. One of the ways to identify which values belong on the X and Y axes is to sketch the chart on paper first. If you can visualize what your chart is supposed to look like, you will have an easier time selecting information correctly and using Excel to construct an effective chart that accurately communicates your message. **Table 4.1** "Key Steps Before Constructing an Excel Chart" provides a brief summary of these points.

Integrity Check

Carefully Select Data When Creating a Chart

Just because you have data in a worksheet does not mean it must all be placed onto a chart. When creating a chart, it is common for only specific data points to be used. To determine what data should be used when creating a chart, you must first identify the message or idea that you want to communicate to an audience.

Table 4.1 Key Steps before Constructing an Excel Chart

Step	Description
Define your message.	Identify the main idea you are trying to communicate to an audience. If there is no main point or important message that can be revealed by a chart, you might want to question the necessity of creating a chart.
Identify the data you need.	Once you have a clear message, identify the data on a worksheet that you will need to construct a chart. In some cases, you may need to create formulas or consolidate items into broader categories.
Select a chart type.	The type of chart you select will depend on the message you are communicating and the data you are using.
Identify the values for the X and Y axes.	After you have selected a chart type, you may find that drawing a sketch is helpful in identifying which values should be on the X and Y axes. In Excel, the axes are: The "category" axis. Usually the horizontal axis – where the labels are found The "value" axis. Usually the vertical axis – where the numbers are found.

Time Series Trend: Line Chart 1

The first chart we will demonstrate is a line chart. Figure 2.61 shows part of the data that will be used to create two line charts. This chart will show the trend of the NASDAQ stock index.

Read more: http://www.investopedia.com/terms/n/nasdaq.asp

This chart will be used to communicate a simple message: to show how the index has performed over a two-year period. We can use this chart in a presentation to show whether stock prices have been increasing, decreasing, or remaining constant over the designated period of time.

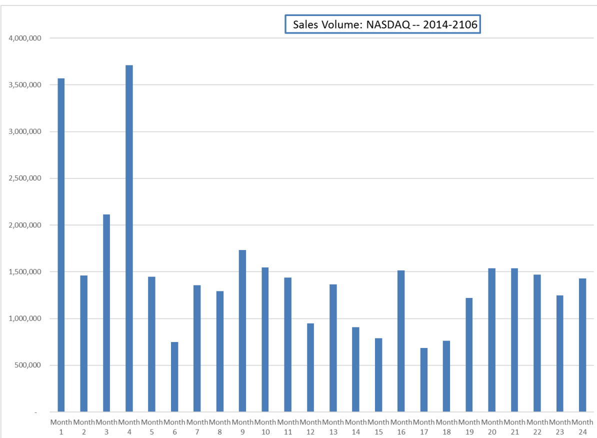


Figure 2.61 Stock Trends

Before we create the line chart, it is important to identify why it is an appropriate chart type given the message we wish to communicate and the data we have. When presenting the trend for any data over a designated period of time, the most commonly used chart types are the line chart and the column chart. With the column chart, you are limited to a certain number of bars or data points. As you increase the number of bars on a column chart, it becomes increasingly difficult to read. As you scroll through the data on the worksheet shown in **Figure 2.61** you will see that there are 24 points of data used to construct the chart. This is generally too many data points to put on a column chart, which is why we are using a line chart. Our line chart will show the volume of sales for the NASDAQ on the Y axis and the Month number on the X axis. The following steps explain how to construct this chart:

If you wish to follow along, download Data file: CH4 Data

1. Open data file *CH4 Data* and save a file to your computer as **CH4 Charting**.
2. Navigate to the **Stock Trend** worksheet.
3. Highlight the range **B4:C28** on the Stock Trend worksheet. (Note – you have selected a label in the first row and more labels in column B. Watch where they show up in your completed chart.)
4. Click the Insert tab of the ribbon.
5. Click the **Line** button in the **Charts** group of commands. Click the first option from the list, which is a basic 2D Line Chart (see **Figure 4.2**).

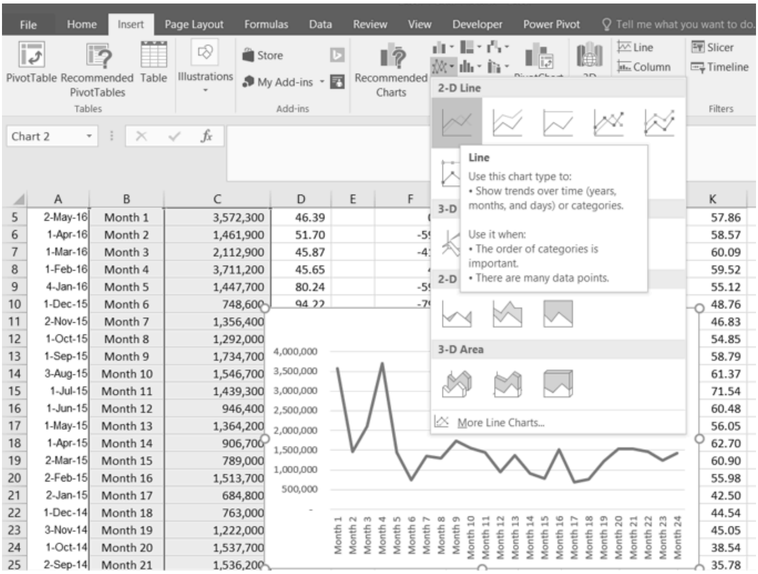


Figure 2.63 Selecting the Basic Line Chart

This adds, or embeds, the line chart to the worksheet, as shown in **Figure 2.63**

Why?

Line Chart vs. Column Chart

We can use both a line chart and a column chart to illustrate a trend over time. However, a line chart is far more effective when there are many periods of time being measured. For example, if we are measuring fifty-two weeks, a column chart would require fifty-two bars. A general rule of thumb is to use a column chart when twenty bars or less are required. A column chart becomes difficult to read as the number of bars exceeds twenty.

Figure 2.63 shows the embedded line chart in the Stock Trend worksheet. Do you see where your labels showed up on the chart?

Notice that additional tabs, or contextual tabs, are added to the ribbon. We will demonstrate the commands in these tabs throughout this chapter. These tabs appear only when the chart is activated.

Note: Excel 2010 uses three contextual tabs for charts. Later versions use only two. Each has all the same tools. They are just organized a little differently.

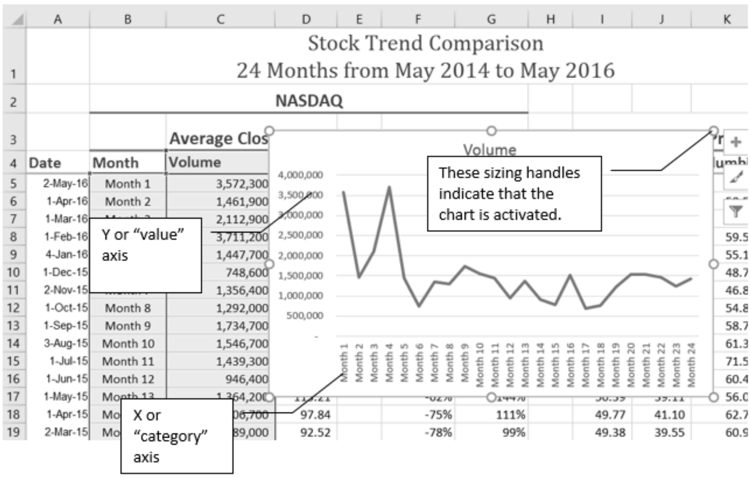
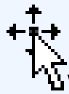


Figure 2.63 Embedded Line Chart in the Stock Trend Worksheet

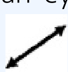
As shown in **Figure 2.63**, the embedded chart is not placed in an ideal location on the worksheet since it is covering several cell locations that contain data. The following steps demonstrate common adjustments that are made when working with embedded charts:

1. **Moving a chart**: Click and drag the upper left corner of the chart to the corner of cell B30.

> Note: Keep an eye on your pointer. It will

change into  when you are in the right place to move your chart.

2. **Resizing a chart**: Place the mouse pointer over the right upper corner sizing handle, hold down the ALT key on your keyboard, and click and drag the chart so it “snaps” to the right side of Column I.

Note: keep an eye on your pointer. It will change into  when you are in the right place to resize your chart

3. Repeat step 2 to resize the chart so the top “snaps” to the top of Row 30, the bottom “snaps” to the bottom of Row 45, and the left side “snaps” to the left side of Column B. Make sure the right side of the chart snaps to the line between column I and J.
4. **Adjusting the chart title**: Click the chart title once. Then click in front of the first letter. You should see a blinking cursor in front of the letter. This allows you to

modify the title of the chart.

5. Type the following in front of the first letter in the chart title: **May 2014-2016 Trend for NASDAQ Sales**.
6. Click anywhere outside of the chart to deactivate it.
7. Save your work.

Figure 2.64 shows the line chart after it is moved and resized. You can also see that the title of the chart has been edited to read **May 2014-2016 Trend for NASDAQ Sales Volume**. Notice that the sizing handles do not appear around the perimeter of the chart. This is because the chart has been deactivated. To activate the chart, click anywhere inside the chart perimeter.

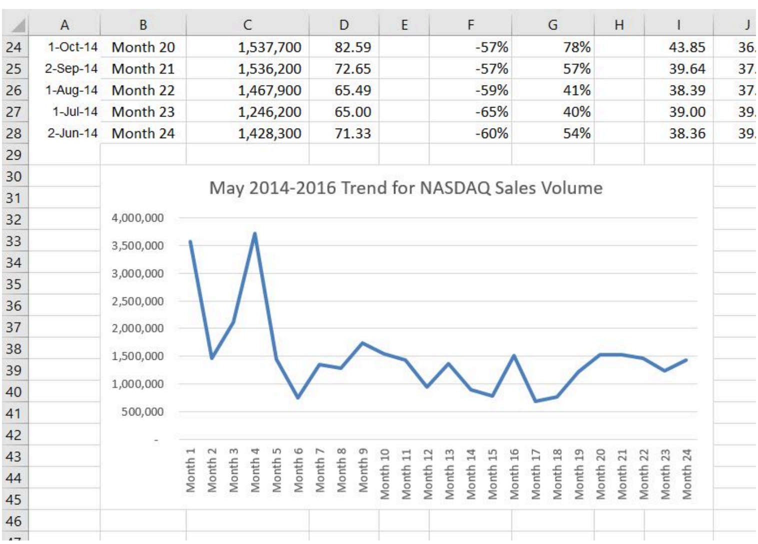


Figure 2.64 Line Chart Moved and Resized

Integrity Check

When using line charts in Excel, keep in mind that anything placed on the X axis is considered a descriptive label, not a numeric value. This is an example of a category axis. This is important because there will never be a change in the spacing of any items placed on the X axis of a line chart. If you need to create a chart using numeric data on the category axis, you will have to modify the chart. We will do that later in the chapter.

Skill Refresher

Inserting a Line Chart

1. Highlight a range of cells that contain data that will be used to create the chart. Be sure to include labels in your selection.

2. Click the Insert tab of the ribbon.
3. Click the Line button in the Charts group.
4. Select a format option from the Line Chart drop-down menu.

Adjusting the Y Axis Scale

After creating an Excel chart, you may find it necessary to adjust the scale of the Y axis. Excel automatically sets the maximum value for the Y axis based on the data used to create the chart. The minimum value is usually set to zero. That is usually a good thing. However, depending on the data you are using to create the chart, setting the minimum value to zero can substantially minimize the graphical presentation of a trend. For example, the trend shown in Figure 4.4 appears to be increasing slightly in recent months. The presentation of this trend can be improved if the minimum value started at 500,000. The following steps explain how to make this adjustment to the Y axis:

1. Click anywhere on the Y (value or vertical) axis on the **May 2014-2016 Trend for NASDAQ Sales Volume** line chart (Stock Trend worksheet).
2. Right Click and select **Format Axis.** The Format Axis

Pane should appear, as shown in **Figure 2.65**.

> Note: If you do not see "Format Axis . . . on your menu, you have not right clicked in the correct spot. Press "Escape" to turn the menu off and try again

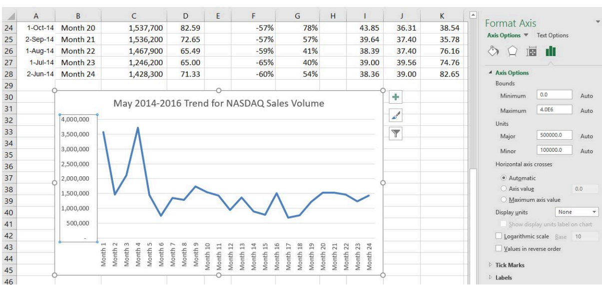


Figure 2.65 Format Axis Pane

1. In the Format Axis Pane, click the input box for the "**Minimum**" axis option and **delete** the zero. Then type the number **500000** and hit Enter. As soon as you make this change, the Y axis on the chart adjusts.
2. Click the X in the upper right corner of the Format Axis pane to close it.
3. Save your work.

Figure 2.66 shows the change in the presentation of the trend line. Notice that with the Y axis starting at 500,000, the trend for the NASDAQ is more pronounced. This adjustment makes it easier for the audience to see the magnitude of the trend.

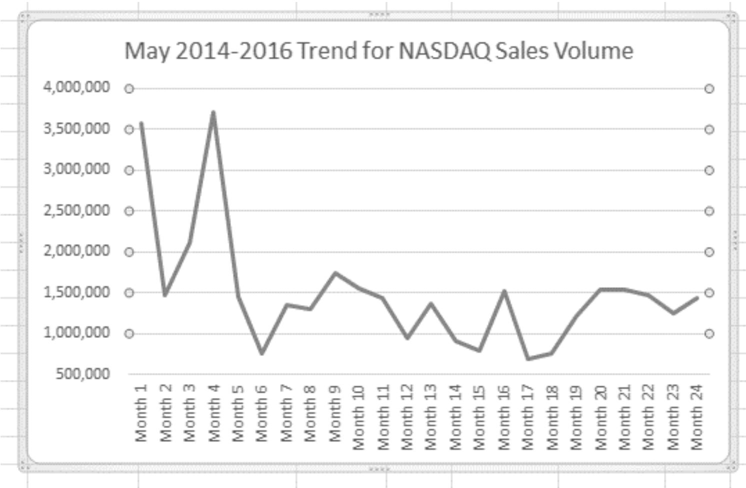


Figure 2.66 Adjusted Y Axis for the S&P 500 Chart

Skill Refresher

Adjusting the Y Axis Scale

1. Click anywhere along the **Y** axis to activate it.
2. Right Click.
   (Note, you can also select the Format tab in the Chart Tools section of the ribbon.)
3. Select **Format Axis** . . .

4. In the **Format Axis pane**, make your changes to the **Axis Options**.
5. Click in the input box next to the desired axis option and then type the new scale value.
6. Click the **Close** button at the top right of the Format Axis pane to close it.

Frequency Distribution: Column Chart

A column chart is commonly used to show trends over time, as long as the data are limited to approximately twenty points or less. A common use for column charts is frequency distributions. A frequency distribution shows the number of occurrences by established categories. For example, a common frequency distribution used in most academic institutions is a grade distribution. A grade distribution shows the number of students that achieve each level of a typical grading scale (A, A−, B+, B, etc.). The Grade Distribution worksheet contains final grades for some hypothetical Excel classes. To show the grade frequency distribution for all the Excel classes in that year, the numbers of students appear on the Y axis and the grade categories appear on the X axis. The number of students for this chart is in Column C. The labels for grades are in

Column A. The following steps explain how to create this chart:

1. Select the Grade Distribution worksheet.
2. Change the years in Row3 to the current academic term and year.
3. Highlight the range **A3:A8** on the **Grade Distribution** worksheet. Column A shows the grade categories.
4. Hold down the **Crtl key**.
5. Without letting go of the **Ctrl key**, select **C3:C8**
6. Click the **Column** button in the **Charts** group section on the **Insert** tab of the ribbon. Select the first option in the 2-D Column section, which is the **Clustered Column** format.
7. Click and drag the chart so the upper left corner is in the middle of cell **H2**.
8. Resize the chart so the left side is locked to the left side of Column H, the right side is locked to the right side of Column O, the top is locked to the top of Row 2, and the bottom is locked to the bottom of Row 16.
9. If Excel displays a legend, delete it by clicking the legend one time and pressing the **DELETE** key on the keyboard. Since the chart presents only one data series, the legend is not necessary.
10. Add the text **Final Grades for** to the chart title. The chart title should now be **Final Grades for All Excel Classes 2016/2017** (or whichever academic year you

are using).

11. Click any cell location on the **Grade Distribution** worksheet to deactivate the chart.
12. Save your work.

Figure 2.67 shows the completed grade frequency distribution chart. By looking at the chart, you can immediately see that the greatest number of students earned a final grade in the B+ to B– range.

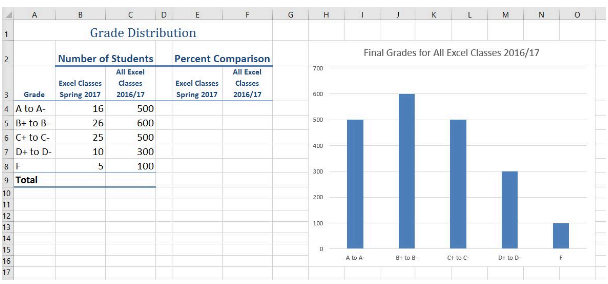


Figure 2.67 Grade Frequency Distribution Chart

Why?

Column Chart vs. Bar Chart

When using charts to show frequency distributions, the difference between a column chart and a bar chart is really a matter of preference. Both are very effective in showing frequency distributions. However, if you are showing a trend over a period of time, a column chart is preferred over a bar chart. This is because a period of time is typically shown horizontally, with the oldest date on the far left and the newest date on the far right. Therefore, the descriptive categories for the chart would have to fall on the horizontal – or category axis, which is the configuration of a column chart. On a bar chart, the descriptive categories are displayed on the vertical axis.

Integrity Check

Too Many Bars on a Column Chart?

Although there is no specific limit for the number of bars you should use on a column chart, a general rule of thumb is twenty bars or less. More is considered a poor use of a column chart because it is difficult to identify meaningful trends or comparisons.

Percent of Total: Pie Chart

The next chart we will demonstrate is a pie chart. A pie chart is used to show a percent of total for a data set at a specific point in time. It is an appropriate chart for qualitative data. The data we will use to demonstrate a pie chart is related to enrollment data for Portland Area Community Colleges for Fall of 2014. You will find that data on the Enrollment Statistics sheet.

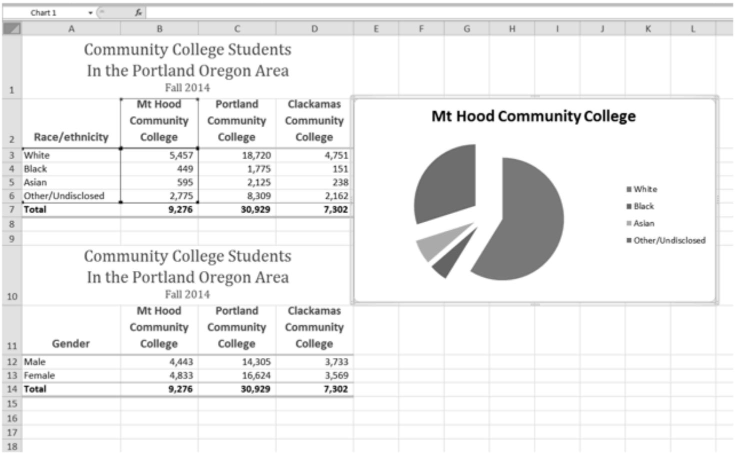
1. Highlight the range **A2:B6** on the **Enrollment Statistics** worksheet.
2. Click the **Insert** tab of the ribbon.
3. Click the **Pie** button in the **Charts** group of commands.
4. Select the first "**2-D Pie**" option from the drop-down list of options.
5. To make the "slices" stand out better, "explode" the pie chart.
    - Click and hold the mouse button down in any of the slices of the pie.
    - Note that you have selection handles on all of the

pie slices.

- Without letting go of your mouse button; drag one of the slices away from the center.
- All of the slices "explode" out from the center.

Note: if you let go of the mouse button before dragging, you may only get one slice to move when you drag it out from the center. This can be another option for displaying your data. Use the Undo button to undo this if you want to try again.

1. Click off the slices and into the white canvas to deselect the pie and select the entire chart.
2. Click and drag the pie chart so the upper left corner is in the middle of cell E2.
3. Resize the pie chart so the left side is locked to the left side of Column E, the right side is locked to the right side of Column L, the top is locked to the top of Row 2, and the bottom is locked to the bottom of Row 10 (see **Figure 2.68**).



Community College Students
In the Portland Oregon Area
Fall 2014

Race/ethnicity	Mt Hood Community College	Portland Community College	Clackamas Community College
White	5,457	18,720	4,751
Black	449	1,775	151
Asian	595	2,125	238
Other/Undisclosed	2,775	8,309	2,162
Total	**9,276**	**30,929**	**7,302**

Community College Students
In the Portland Oregon Area
Fall 2014

Gender	Mt Hood Community College	Portland Community College	Clackamas Community College
Male	4,443	14,305	3,733
Female	4,833	16,624	3,569
Total	**9,276**	**30,929**	**7,302**

Figure 2.68 Pie Chart Moved and Resized

1. Click the chart **legend** once and press the **DELETE** key on your keyboard. A pie chart typically shows labels next to each slice. Therefore, the legend is not needed.
2. Right click any of the slices in the pie chart, and select **Add Data Labels** from the list. This will add the values for each of the slices in the pie.
3. Now, you can right click one of the numbers and select **Format Data Labels** from the list. This will open the **Format Data Labels** pane on the right.
4. Check the boxes for **Category Name** and **Percentage** in the **Label Options** section in the **Format Data Labels** pane. This will add the Race/ethnicity labels as well as the percentage data to the pie chart.

5. Uncheck the box next to the **Value** box. This will remove the numbers from the pie chart (see **Figure 2.69**).
6. Click the **Close** button at the top of the Format Data Labels pane.
7. Select the data labels again (if needed). Click the **Home** tab of the ribbon and then click the **Bold** button. This will bold the data labels on the pie chart.
8. Save your work.

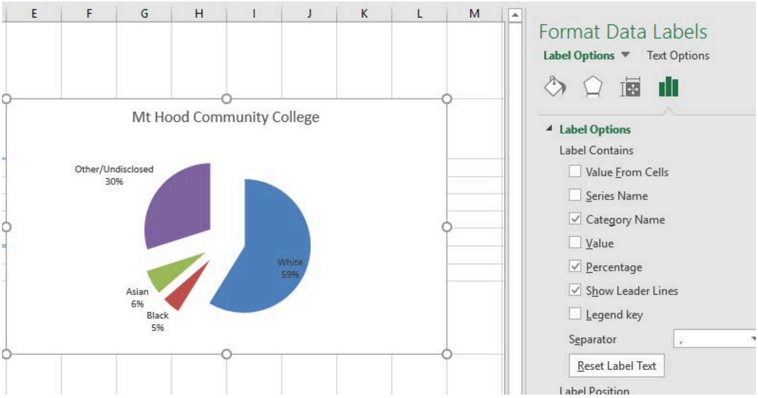
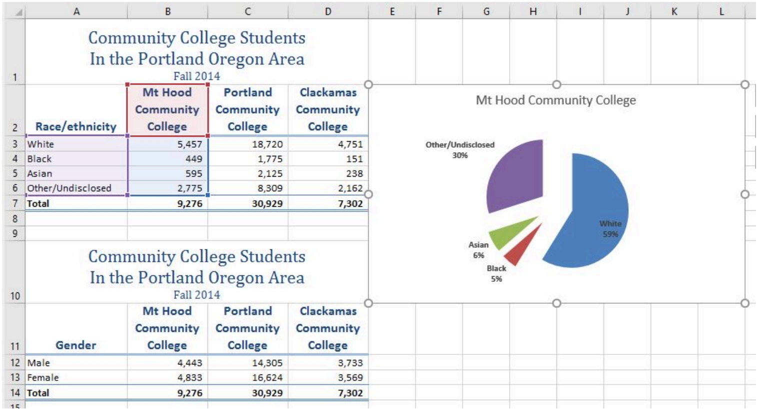


Figure 2.69 Final Settings in the Format Data Labels Pane

Although there are no specific limits for the number of categories you can use on a pie chart, a good rule of thumb is ten or less. As the number of categories exceeds ten, it becomes more difficult to identify key categories that make up the majority of the total.



Community College Students
In the Portland Oregon Area
Fall 2014

Race/ethnicity	Mt Hood Community College	Portland Community College	Clackamas Community College
White	5,457	18,720	4,751
Black	449	1,775	151
Asian	595	2,125	238
Other/Undisclosed	2,775	8,309	2,162
Total	9,276	30,929	7,302

Community College Students
In the Portland Oregon Area
Fall 2014

Gender	Mt Hood Community College	Portland Community College	Clackamas Community College
Male	4,443	14,305	3,733
Female	4,833	16,624	3,569
Total	9,276	30,929	7,302

Figure 2.610 Final Enrollment Statistics Pie Chart

Skill Refresher

Inserting a Pie Chart

1. Highlight a range of cells that contain the data you will use to create the chart.
2. Click the Insert tab of the ribbon.
3. Click the Pie button in the Charts group.
4. Select a format option from the Pie Chart drop-down menu.

Percent of Total: Stacked Column Chart

You will not have to construct this chart in our course, but reading through how it is constructed will help you when we begin our statistics unit. We use a stacked column chart to show a percent of a total . For example, the data on the Enrollment Statistics worksheet shows student enrollment by race for several colleges. We would like to see all of the data on all of the colleges.

1. Highlight the range **A2:D6** on the **Enrollment Statistics** worksheet.
2. Click the **Insert** tab of the ribbon.
3. Click the **Column** button in the **Charts** group of commands. Select the **100% Stacked Column** format option from 2-D Column section in the drop-down list (see **Figure 2.611**).

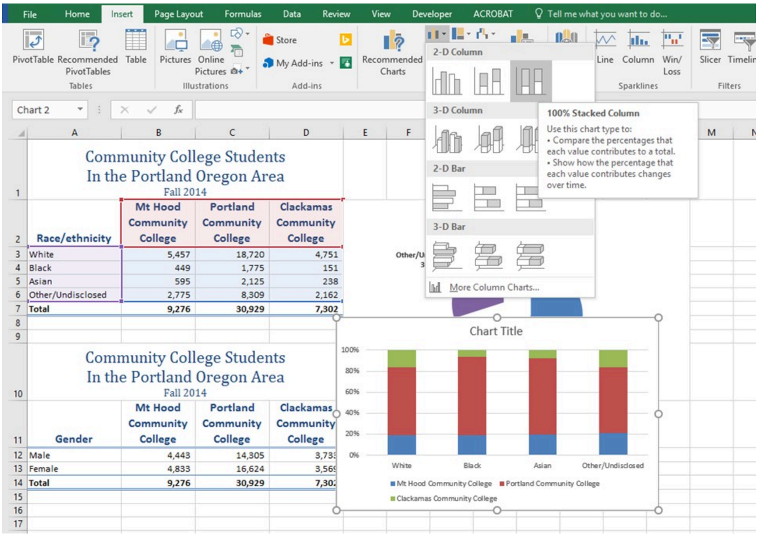


Figure 2.611 Selecting the 100% Stacked Column Chart

Figure 2.612 shows the column chart that is created after selecting the 100% Stacked Column format option. As mentioned, the goal of this chart is to show the enrollment of students by race. However, notice that Excel places the racial categories on the X axis. It would be more useful if the different colleges were there instead.

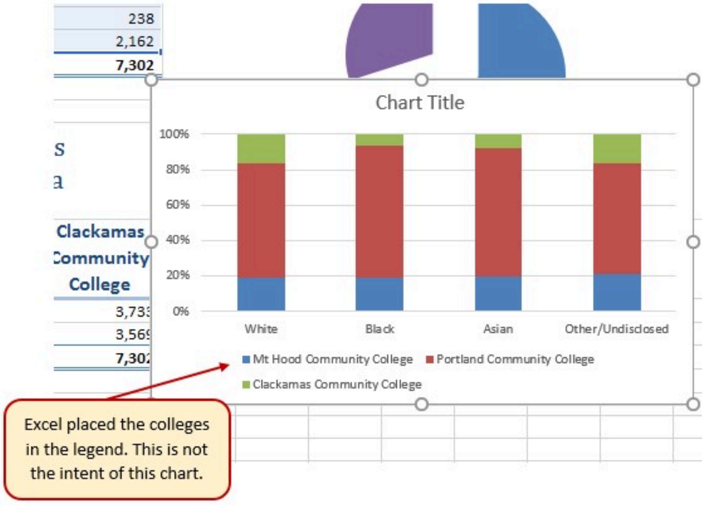


Figure 2.612 Initial Construction of the 100% Stacked Column Chart

The reason that Excel organized the data this way is that there are more Race/ethnicity categories (data in column A) than there are colleges (data in row 2). Not a bad guess. But, not what we wanted in this case.

The remaining steps explain how to correct this problem and complete the chart:

1. Click the **Switch Row/Column** button in the **Design** tab on the **Chart Tools** section of the ribbon. This reverses the legend and current X axis categories.

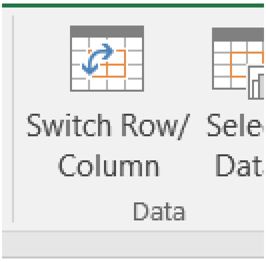


Figure 4.24 Switch Row/Column.

2. Click and drag the chart so the upper left corner is in the middle of cell **E12**.
3. Resize the chart so the left side is locked to the left side of Column E, the right side is locked to the right side of Column N, the top is locked to the top of Row 12, and the bottom is locked to the bottom of Row 30.
4. Click the **legend** one time and press the **DELETE** key on your keyboard.
5. **Add a Data Table**. This is another way of displaying a legend for a column chart along with the numerical values that make up each component.

    - In earlier versions of Excel, find the **Labels** group of commands and select the **Show Data Table with Legend Keys** option from the drop-down menu.
    - In Excel 2016, find the **Add Chart Element** tool on the **Design** tab, select **Data Table With**

Legend Keys

6. Change the Chart Title to **Enrollment by Race**.
    - If there is no chart title, you will need to add one using the **Add Chart Element** tool on the **Design** tab.
7. Save your work.

Figure 2.613 shows the final stacked column chart. Notice the similarities and differences in the enrollment at the local community colleges.

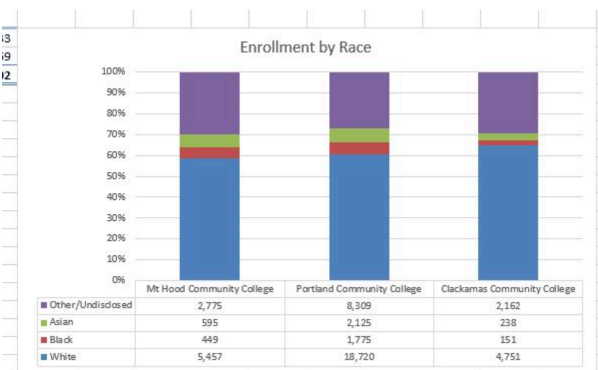


Figure 2.613 Final 100% Stacked Column Chart

Skill Refresher:

Inserting a Stacked Column Chart

1. Highlight a range of cells that contain data that will be used to create the chart.
2. Click the Insert tab of the ribbon.
3. Click the Column button in the Charts group.
4. Select the Stacked Column format option from the Column Chart drop-down menu to show the values of each category on the Y axis. Select the 100% Stacked Column option to show the percent of total for each category on the Y axis.

Representing Quantitative Data: the Histogram

This is the last chart we will cover in this chapter, and perhaps the one we will use most in our class. Histograms are used to display quantitative data. For example, the data on this math test is displayed in a frequency distribution and a histogram in **Figure 2.614**. Notice that the bins are ranges of numbers, not attributes:

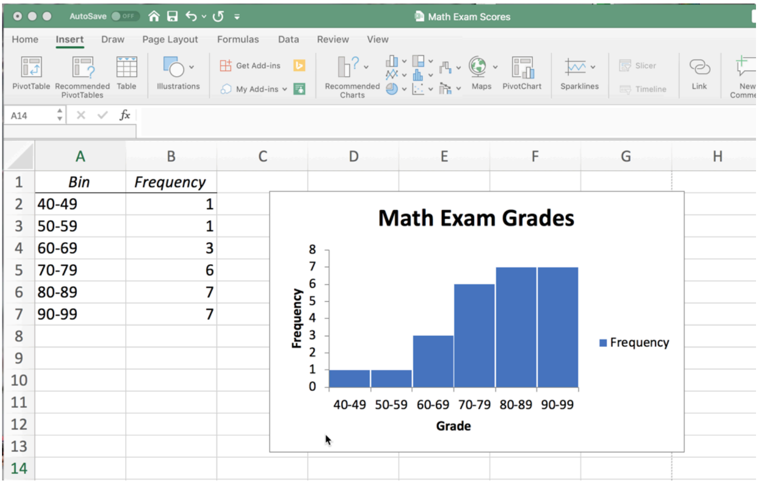


Figure 2.614 Histogram of Math Grades

While the directions for creating a histogram vary due to the version of Excel you have, these directions work for most versions:

1. Use **Max** and **Min** functions to determine the class sizes of this data.
2. Set up class sizes, and then create a new 'Bin' column to record the upper limit of each class.
3. Click on 'Data' to see if 'Data Analysis' is in the upper right corner of the ribbon. If it is not, you will need to upload the Data Analysis Toolpak (directions are in the video).
4. Click on 'Data Analysis' and select 'Histogram.'

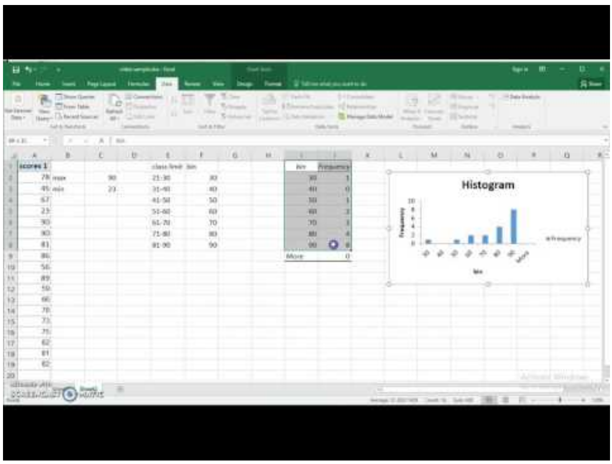
5. A pop-up window will appear. You want to include the range of your data in 'Input Range.' To do that, select your data and the range of numbers will appear in that section of the pop-up.
6. For 'Bin Range' include the column of upper class limits.
7. Enable 'Output Range' and select the cell on your sheet where you would like the histogram and the frequency distribution to appear.
8. Click on 'Chart Output' so that the histogram will be pasted on this page.

Notice the histogram will appear with gaps between the bars and only the upper limit of each class showing. You will want to edit your chart:

1. Use the original class limits and frequency to create a well-formatted frequency chart (if needed).
2. Copy the original class limits into the 'Bin' in the created frequency chart to change the horizontal labels on your histogram.
3. Right click on any bar, and then select 'Format Data Point.' You will see a slider for gap width. Reduce the gap width to zero or less than 5% if you want a bit of a gap.
4. Replace or copy the entire range of each class size into the 'bin' in the frequency chart.

5. See Section 2.8 of this chapter for more formatting instructions.

This video explains the process in more detail:



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=723

There is also a histogram option under 'Insert' in the Excel ribbon. Highlight your data first and then select this option. You cannot set the class limits, but you can adjust the number of bins and the width of the bins by right-clicking or double-clicking on the bars of the created histogram. You can also add labels by clicking once on the histogram chart and then selecting 'Chart Design' tab. If you do not mind where the classes start numerically, this may be a viable option!

Skill Refresher:

Inserting a Histogram

1. Click on the Data tab of the ribbon to be sure that the Data Analysis Toolpak has been installed. If not, follow the directions above to install this add-in.
2. Be sure you have set up a column of the upper range of each class in your histogram.
3. Click on Data Analysis in the upper right corner of the ribbon and select histogram.
4. Select the data you want to be analyzed (input range).
5. Click on bin range and then select the upper range of values from Step 2.
6. Check off Output Range and select a cell from the area where you would like the data to display.
7. Check of Chart Output and then click OK.

Key Takeaways

- Identifying the message you wish to convey to an audience is a critical first step in creating an Excel chart.
- Both a column chart and a line chart can be used to present a trend over a period of time. However, a line chart is preferred over a column chart when presenting data over long periods of time.
- The number of bars on a column chart should be limited to approximately twenty bars or less.
- When working with frequency distributions, the use of a column chart or a bar chart is a matter of preference. However, a column chart is preferred when working with a trend over a period of time.
- A pie chart is used to present the percent of total for a categorical (or qualitative) data set.
- A stacked column chart is used to show how a percent total changes over time.
- A histogram is used to represent quantitative data.

Attribution

Adapted by Noreen Brown from How to Use Microsoft Excel: The Careers in Practice Series, adapted by The Saylor Foundation without attribution as requested by the work's original creator or licensee, and licensed under CC BY-NC-SA 3.0.

Media Attributions

- fig4-2
- fig4-3
- Figure 4-4
- Figure 4-5
- Figure 4-12
- fig19
- Figure 4.20 Final Settings in the Format Data Labels Pane
- Figure 4-21
- Figure 4-22
- Figure 4-23
- fig
- Figure 4-24
- Math 40

FORMATTING CHARTS

Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker

Learning Objectives

1. Apply formatting commands to the X and Y axes.
2. Enhance the visual appearance of the chart title and chart legend by using various formatting techniques.
3. Assign titles to the X and Y axes that clarify labels and numeric values for the reader.
4. Apply labels and formatting techniques to the data series in the plot area of a chart.
5. Apply formatting commands to the chart area and the plot area of a chart.
6. Employ series lines and annotations to enhance trends and provide additional information on a chart.

You can use a variety of formatting techniques to enhance the appearance of a chart once you have created it. Formatting commands are applied to a chart for the same reason they are applied to a worksheet: they make the chart easier to read. However, formatting techniques also help you qualify and explain the data in a chart. For example, you can add footnotes explaining the data source as well as notes that clarify the type of numbers being presented (i.e., if the numbers in a chart are truncated, you can state whether they are in thousands, millions, etc.). These notes are also helpful in answering questions if you are using charts in a live presentation. We will demonstrate these formatting techniques using the column chart and stacked column chart from the previous section.

X and Y Axis Formats

There are numerous formatting commands we can apply to the X and Y axes of a chart. Although adjusting the font size, style, and color are common, many more options are available through the Format Axis pane. The following steps demonstrate a few of these formatting techniques on the **Grade Distribution Comparison** chart:

1. Switch to the **Grade Distribution** worksheet and click anywhere along the X axis (horizontal axis) of the Grade Distribution Comparison chart.
2. Right click and select Font.
3. Change the font to Arial, the Font Style to Bold, and the

Size to 11 (see **Figure 2.71**).

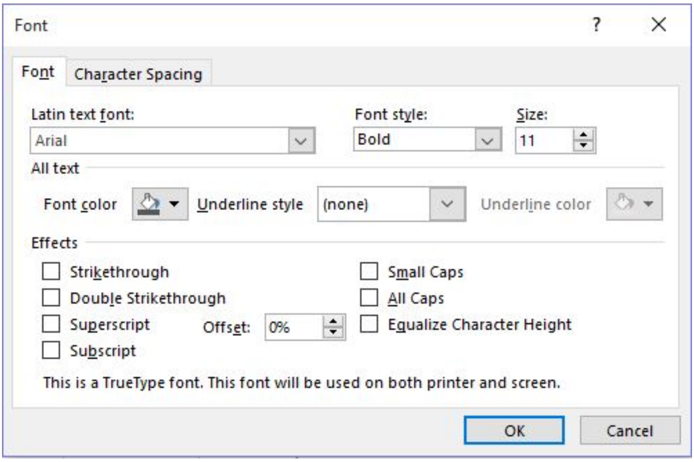


Figure 2.71 Font Dialog Box

4. Click anywhere along the Y axis to activate it and repeat steps 2 and 3.
5. Click on the chart title and repeat steps 2 and 3, but set the Size to 14.
6. The final appearance of the axes is shown in Figure 2.72 **Formatted X & Y Axes.**

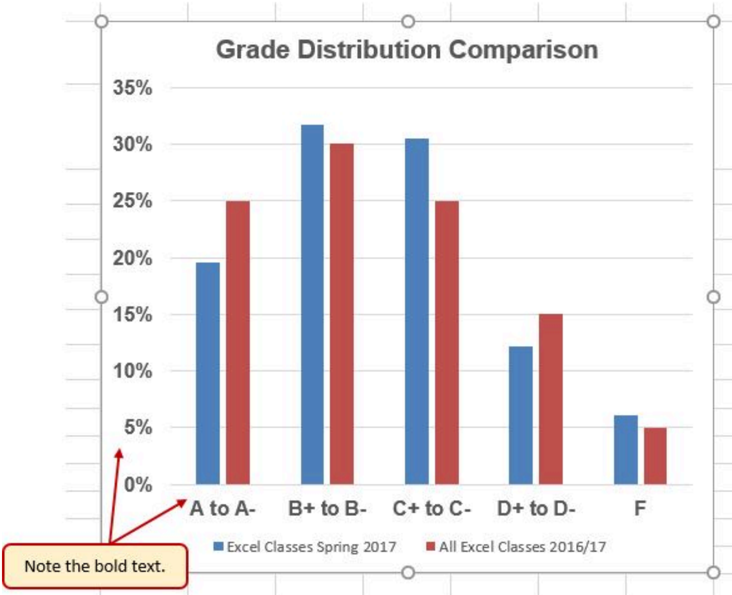


Figure 2.72 Formatted X and Y Axes

Next we want to make some changes to the percentage numbers on the Y (vertical) axis.

1. Right click the vertical (value) axis. Select **Format Axis**. This opens the **Format Axis** pane.
2. Click **Number** from the list of options. The commands in this section of the **Format Axis** pane are used to format numbers that appear on the selected axis of the chart.
3. Click in the Decimal places input box and change the value to **1.**

4. Select **Axis Options**. Change the **Minimum Bound** to **.05** to make the differences in the columns more dramatic. The Format Axis pane should match **Figure 2.73**.
5. Click the **Close** button at the top of the Format Axis pane.
6. Save your work.

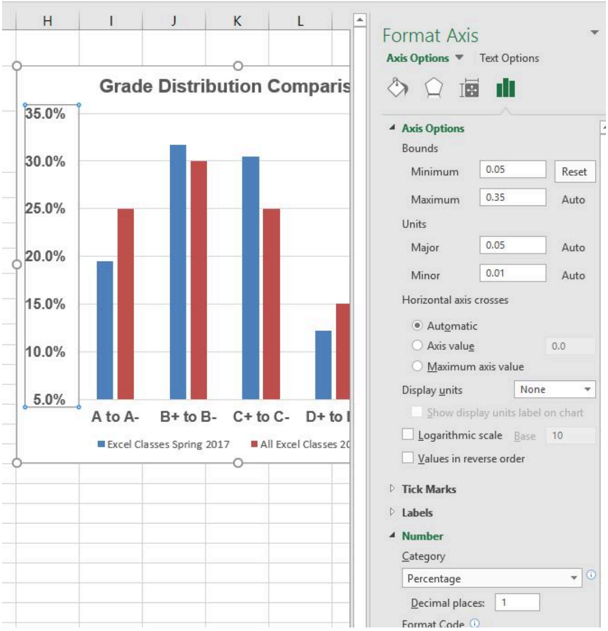


Figure 2.73 Format Axis Pane changes

Note: Experiment! You can also change font styling using shortcut keys and the buttons on the **Home tab.**

Skill Refresher

Formatting the X and Y Axes

1. Click anywhere along the X or Y axis to activate it.
2. Click either the Home tab or Design tab of the ribbon.
3. Select any of the available formatting commands in these tabs.

Skill Refresher

X and Y Axis Number Formats

1. Click anywhere along the X or Y axis to activate it.
2. Click the Layout tab in the Chart Tools section of the ribbon.
3. Click the Format Selection button in the Current Selection group of commands.
4. Click Number from the list of options on the left side of the Format Axis dialog box.
5. Select a number format and set decimal places on the right side of the Format Axis dialog box.
6. Click the Close button in the Format Axis pane.

Chart Legend and Title Formats

The next items we will format on the Grade Distribution Comparison chart are the chart legend and title. Similar to the how we formatted the X and Y axes, we can format these items by activating them and using the formatting commands in the Home tab or the Format pane. The following steps explain how to add these formats:

1. Right click the legend on the **Grade Distribution Comparison** chart and select **Format Legend**.
2. Select **Right** in the Legend Position options. Close the

Format Legend pane.

3. Move the legend by placing your cursor – shaped like a little plus sign with four arrows – on the edge of the selection box. Click and drag the legend so the top of the legend aligns with the 35% line next to the plot area (see **Figure 2.74**).

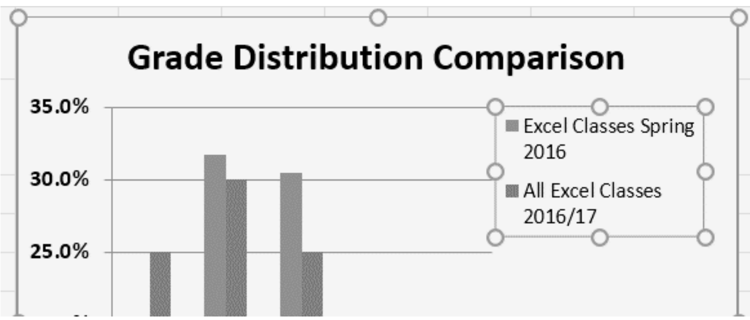


Figure 2.74 Moving the Legend

1. While the legend is still selected, change the font style in the Home tab of the ribbon to Arial.
2. Change the font size to 12 points.
3. Click the bold and italics commands in the Home tab of the ribbon.
4. Click and drag the left sizing handle so the legend is against the plot area (see **Figure 2.75**).

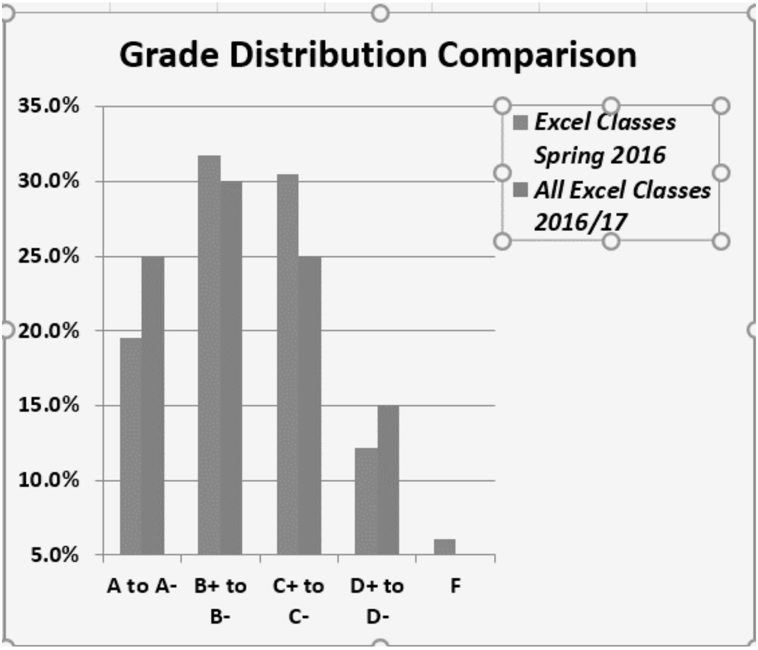


Figure 2.75 Legend Formatted and Resized

1. Click the chart title to activate it.
2. Right click on the chart title and select **Format Chart Title** to open the **Format Chart Title** pane.
3. Under **Title Options** , in the **Effects** group (the option in the middle) give your title one of the **Preset** shadows. Change the color, if you like.
4. Close the Format Chart Title pane.
5. Save your work.

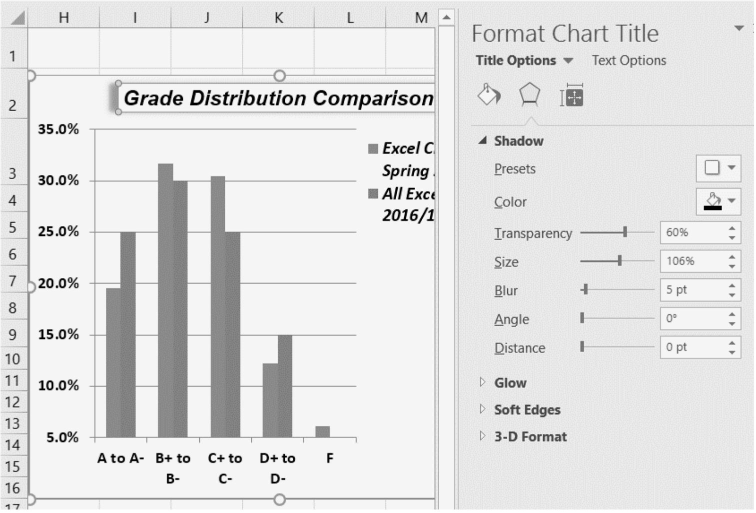


Figure 2.76 Format Chart Title Pane

Skill Refresher

Formatting the Chart Legend

1. Click the Legend to activate it.
2. Click either the Home tab or right click to activate the appropriate formatting pane.
3. Select any of the available formatting commands.
4. Click and drag the legend to move it.
5. Click and drag any of the sizing handles to

adjust the size of the legend.

Skill Refresher

Formatting the Chart Title

1. Click anywhere on the chart title.
2. Click either the Home tab or right click to activate the appropriate formatting pane.
3. Select any of the available formatting commands.

X and Y Axis Titles

Titles for the X and Y axes are necessary for defining the numbers and categories presented on a chart. For example, by looking at the Grade Distribution Comparison chart, it is not clear what the percentages along the Y axis represent. The following steps explain how to add titles to the X and Y axes to define these numbers and categories:

1. Click anywhere on the Grade Distribution Comparison chart in the Grade Distribution worksheet to activate it.
2. On the **Design** tab on the ribbon select the Add Chart Element button, then Axis Titles, then Primary Vertical. (See **Figure 2.77**)

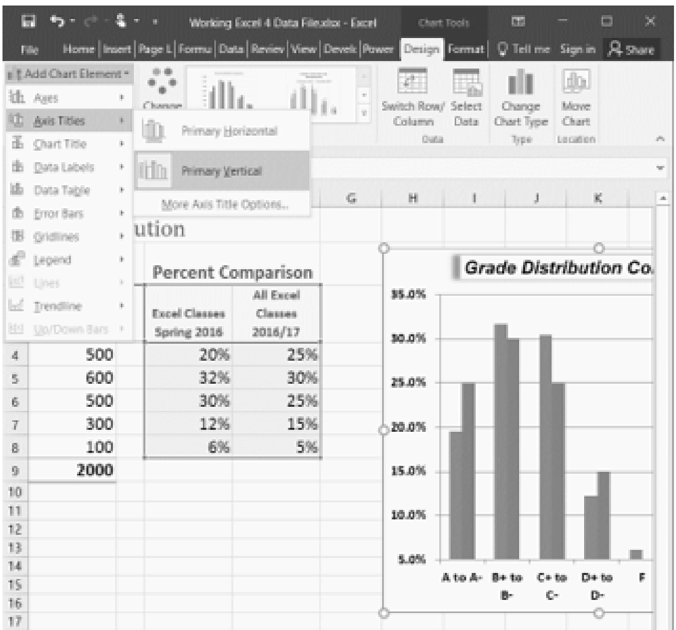


Figure 2.77 Selecting a Title for the Y Axis

3. Using the Home ribbon, change the font of the axis title to Arial, Bold, size 11.
4. Click in the beginning of the Y axis title and delete the generic title. Type Percent of Enrolled Excel

Students.(see **Figure 2.78**).

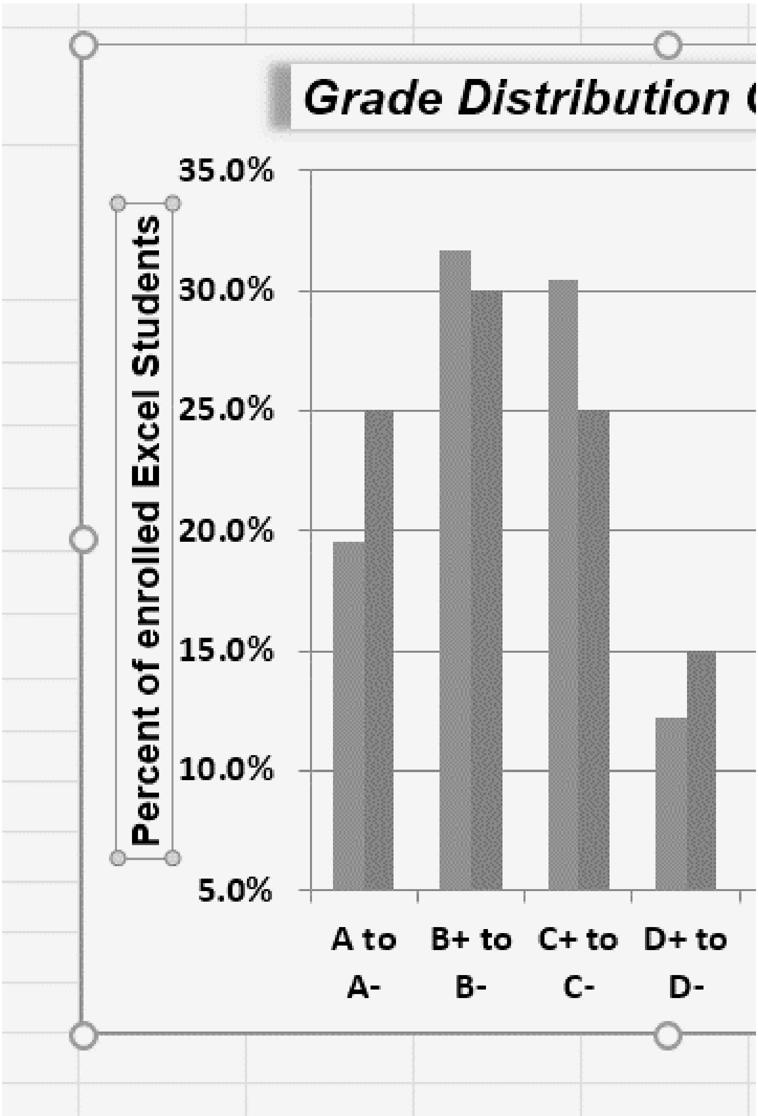


Figure 2.78 Adding and Formatting the Y Axis Title

Next we will add the title for the X axis.

1. On the **Design** tab select the **Add Chart Element** button, then **Axis Titles**, then **Primary Horizontal**.
2. Using the Home ribbon, change the font of the axis title to Arial, Bold, size 11.
3. Click in the beginning of the X axis title and delete the generic title. Type **Final Course Grade**. **Figure 2.79** shows the added titles for the X and Y axes. The titles provide definitions for the grade categories along the X axis as well as the percentages on the Y axis.
4. Save your work.

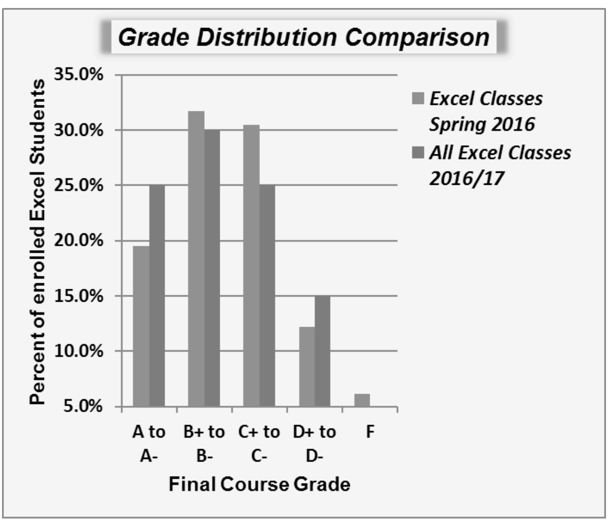


Figure 2.79 X and Y Axis Titles Added

Skill Refresher

X and Y Axis Titles

1. On the **Design tab** select the **Add Chart Element** button.
2. Click anywhere on the chart to activate it.
3. Select one of the options from the second drop-down list.
4. Click in the axis title to remove the generic title and type a new title.

Data Series Labels and Formats

Adding labels to the data series of a chart is a key formatting feature. A data series is the item that is being displayed graphically on a chart. For example, the blue bars on the Grade Distribution Comparison chart represent one data series. We can add labels at the end of each bar to show the exact percentage the bar represents. In addition, we can add other formatting enhancements to the data series, such as changing the color of the bars or adding an effect. The following steps explain how to add these labels and formats to the chart:

1. Click on any of the the red bars representing the **All Excel Classes** data series on the **Grade Distribution Comparison** chart in the **Grade Distribution** worksheet. Clicking one bar automatically activates all bars in the data series. If you click a bar a second time, only that bar is activated.
2. Right click and select **Format Data Series** to open up the **Format Data Series** pane.
3. Click the **Fill and Line** (paint bucket) button to bring up the Fill and Border group of commands.
4. Click the word **Fill** (if needed) to expand the list of Fill options.
5. Select **Pattern Fill**. Then select **30%** (fifth column, top row). Changing your fill pattern to a pattern makes it easier to distinguish between the data series when you print or view your chart in black and white. While you are there, make changes to the fill by experimenting with different foreground and background colors.
6. Close the **Format Data Series** pane.

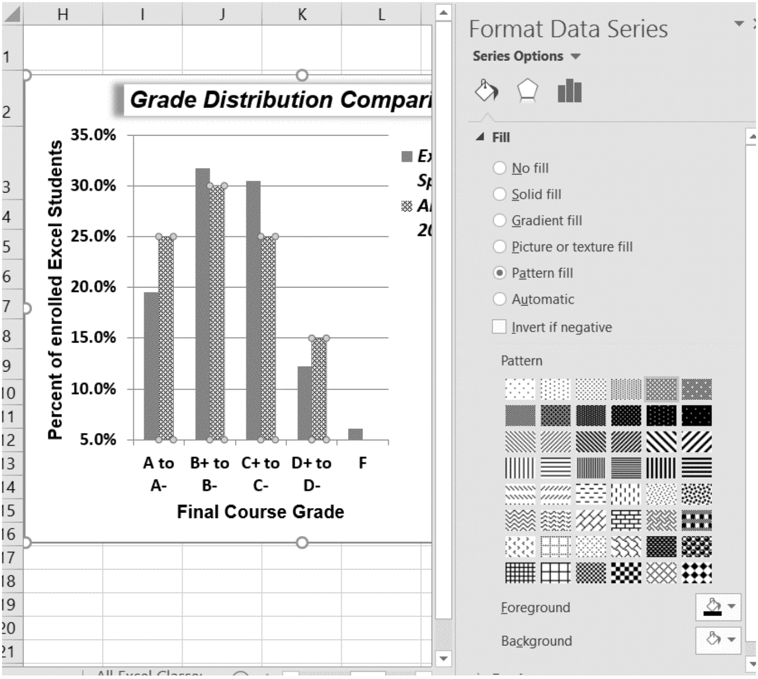


Figure 2.710 Changing the Fill of a Data Series

Now we are going to add the Data Labels at the end of the columns.

1. Be sure that your entire chart is selected, not just one of the data series. Click the **Design** tab in the **Chart Tools** section of the ribbon.
2. On the **Design** tab select the **Add Chart Element** button, then **Data Labels**, then **Outside End** (see **Figure 2.711**)
3. Click on one of the **Data Labels**. Note that all of the

data labels for that data series are selected.

4. Using the **Home** ribbon, change the font to Arial, Bold, size 9.
5. Click on one of the data labels for the other data series. Format those data labels as Arial, Bold, size 9 as well.
6. Save your work.

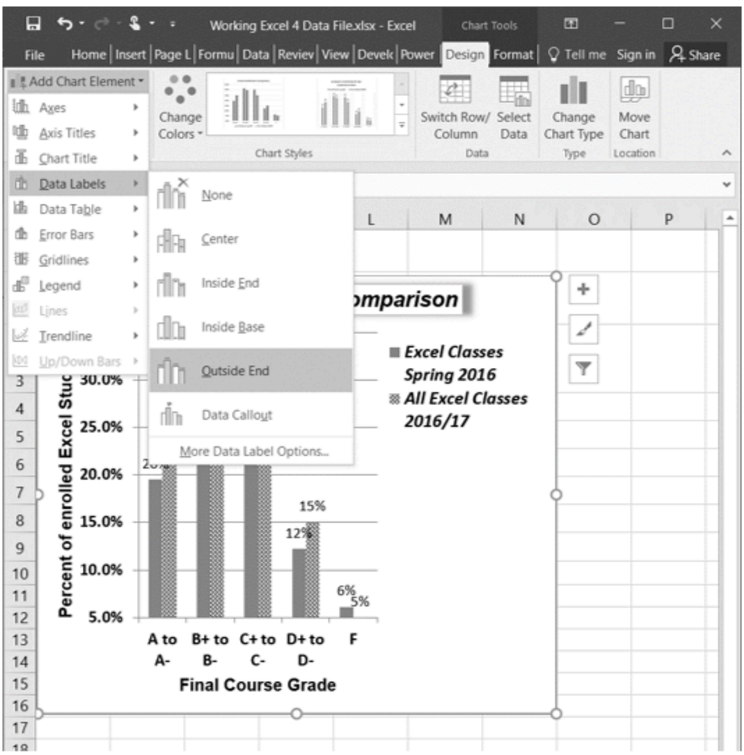


Figure 2.711 Adding Labels to a Data Series

Figure 2.712 shows the Grade Distribution Comparison chart with the completed formatting adjustments and labels added to the data series. Note that we can move each individual data label. This might be necessary if two data labels overlap or if a data label falls in the middle of a grid line. To move an individual data label, click it twice, then click and drag.

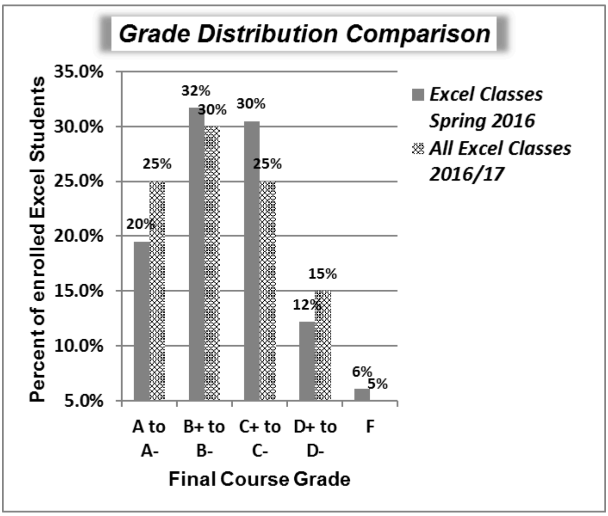


Figure 2.712 Completed Formatting Adjustments for the Data Series

Skill Refresher:

Adding Data Labels

1. Click anywhere on the chart to activate it.
2. Click the Design tab in the Chart Tools section

of the ribbon.

3. Click the **Add Chart Element** in the **Chart Layout** group.
4. Then, select **Data Labels**
5. Select one of the preset positions from the drop-down list.

Skill Refresher

Formatting a Data Series

1. Click any bar or line for a data series.
2. Right click to activate the **Format Data Series** pane.
3. Use the formatting tools in the pane to make changes to the data series.

Adding Series Lines and Annotations to a Chart

The last formatting features we will demonstrate are adding series lines and annotations to a chart. To demonstrate these skills, we will use the Change in Enrollment Statistics Spend Source stacked column chart. Series lines are commonly used in stacked column charts to show the change from one stack to the next. Annotations are useful for clarifying the data presented in a chart or for identifying data sources. In addition to demonstrating these skills, we will review several of the formatting skills that were covered in this section. The following steps include the skills review as well as the new formatting features:

1. Locate the **Enrollment by Race** stacked column chart on the **Enrollment Statistics** worksheet. Activate the chart by clicking anywhere inside the chart perimeter.
2. Move the chart to a separate chart sheet by clicking the **Move Chart** button in the **Design** tab of the ribbon. Type the following in the New sheet input box: **Enrollment by Race Chart**. Click the **OK** button.
3. Click anywhere on the data table (on the x axis) to activate it. Using the **Home** ribbon, change the font to Arial, Bold, size 12.
4. Activate the **Y** axis and apply the same formatting adjustments as stated in step 3.

5. Add a **Y** axis title using **Add Chart Elements** – **Axis Titles** – then **More Axis Title Options**.
6. In the **Format Axis Title** pane change the fill color and border to colors of your choice.
7. Then, using the **Home** tab of the ribbon, change the font to Arial, Bold, size 14.
8. Change the text of the **Y** axis title to **Percent Enrollment by Race**.
9. Check the horizontal axis to see if this process created an extra axis title there. If it did, delete it.
10. Activate the title of the chart by clicking it once. The **Format Chart Title** pane should be open. If not, right click the Chart title and select **Format Chart Title** from the menu. Change the fill and border to match your vertical Axis label.
11. Then, using the **Home** tab of the ribbon, change the font of the chart title to Arial, Bold, size 20.
12. Close the **Formatting** pane.
13. Click the **Add Chart Elements** tool (Design tab), then **Lines**, then **Series Lines**.

    This adds lines to the chart, connecting each data series between the three stacks (see **Figure 2.713**).

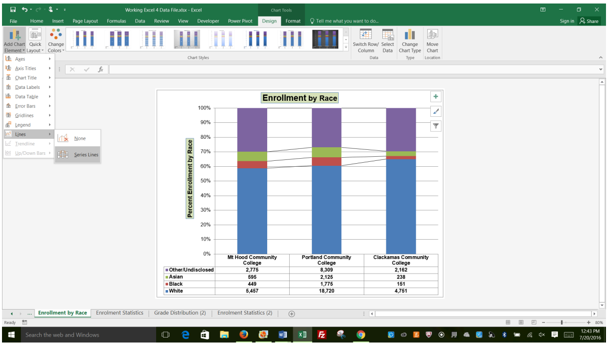


Figure 2.713 Selecting the Series Lines Option

1. Right click on any of the series lines added to the chart. Clicking one line will activate all lines on the chart. (If the Format Pane is open, you will not need to right click. Just left click on any of the series lines to change the format pane to Format Series Lines)
2. Select **Format Series Lines**. This will open the **Format Series Lines** pane.
3. Change the width to **2.25**.
4. Close the **Format Series Lines** pane.

Figure 2.714 shows the appearance of the chart with the series lines connecting the two stacks. This formatting enhancement is common for stacked column charts. The lines help focus the audience's attention to changes in the percent of total trend.

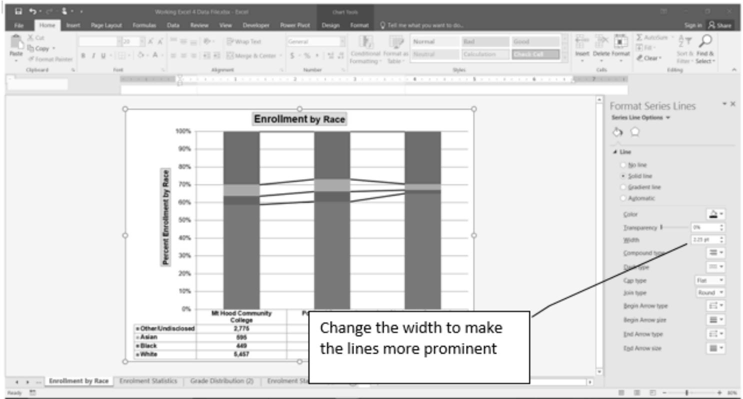


Figure 2.714 Series Lines Added to the Stacked Column Chart

Our chart demonstrates the percentage differences in enrollment between the community colleges. But, it would be handy to know the total Enrollment at each of the colleges. To display that, we will add text boxes above each column. To start with, we need to make room for the text boxes.

1. Select the **Plot Area**. Place your cursor on the top center handle of the Plot Area and drag down about ½ inch.

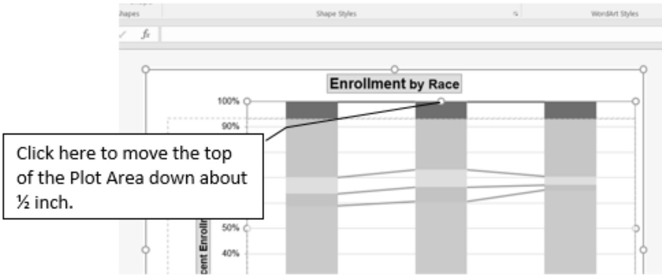


Figure 2.715 Resizing the Plot Area

Add text boxes to include additional information in the chart.

1. Click the **Text Box** button in the **Text** group on the **Insert** tab of the ribbon (see **Figure 4.41**).
2. Place the mouse pointer on the left edge of the chart area approximately one-quarter inch from the top. Click and drag a rectangle approximately one and a half inches wide and one-quarter inch high (see **Figure 2.716**). Don't worry if it's not exact – you can move and resize text boxes at any time.
3. Type **Total Enrollment**. This tells the audience the size of each school.
4. Select all of the text in the text box. (You can either highlight the text or click on the border of the text box

once to select all of the text). Using the **Home** tab of the ribbon, change the font to Arial, size 14.

5. Repeat the process to add and format text boxes above each column. You can try to copy and paste the text boxes if you would like to save time.
6. In each text box, type the Total Enrollment for each school:
    - Mt Hood – 9,276
    - Portland – 30,929
    - Clackamas – 7,302
7. Save your work.

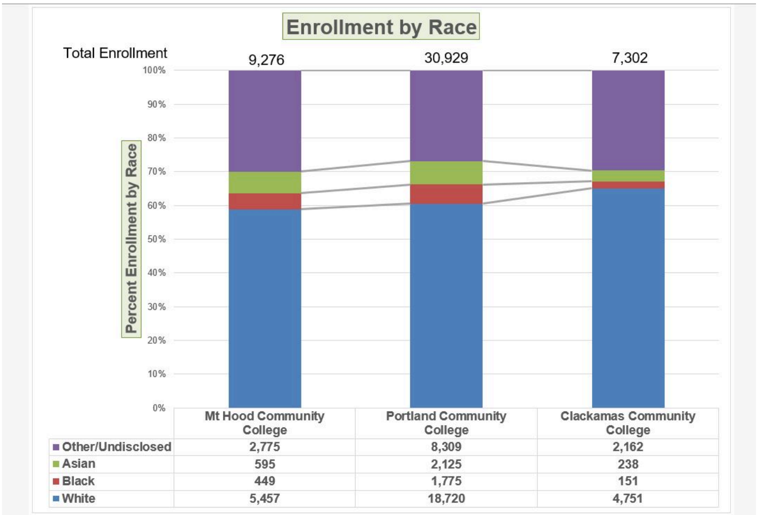


Figure 2.716 Completed Stacked Column Chart

Integrity Check

Annotations and Axis Titles

Although adding annotations and axis titles can be a tedious process, doing so maintains a high level of integrity for your charts. People can misinterpret the message being conveyed by the chart if they make inaccurate assumptions about the values displayed. Axis titles and annotations help prevent readers from making false assumptions and ensure that readers see the most accurate representation of the message being conveyed by the chart.

Skill Refresher

Adding Series Lines

1. Click anywhere on the chart area.

2. Click the Layout tab of the ribbon.
3. Click the Lines button in the Analysis group of commands.
4. Click the Series Lines option from the drop-down list.

Skill Refresher:

Adding Annotations

1. Click anywhere on the chart area.
2. Click the Insert tab of the ribbon.
3. Click the Text Box button in the Text group of commands.
4. Click and drag the size of the text box needed on the chart.
5. Apply any desired format changes from the Home tab of the ribbon.
6. Type the desired text.

Key Takeaways

- Applying appropriate formatting techniques is critical for making a chart easier to read.
- Many formatting commands in the Home tab of the ribbon can be applied to a chart.
- To change the number format for a data label, you must use the Number section in the Format Data Labels dialog box. You cannot use the Number format commands in the Home tab of the ribbon.
- To change the number format for the values on the Y axis, and the X axis in the case of a scatter chart, you must use the Number section of the Format Axis dialog box. You cannot use the Number format commands in the Home tab of the ribbon.
- Axis titles and annotations help prevent false assumptions from being made and ensure that the reader sees the most accurate representation of the information presented on a chart.

Attribution

Media Attributions

- Figure 4-26
- Figure 4-27
- Figure 4-28
- figure-4-30-selecting-a-title-for-the-y-axis
- figure-4-34-adding-labels-to-a-data-series
- figure-4-42-series-lines-added-to-the-stacked-column-chart
- fig40
- Figure 4-41

USING CHARTS WITH MICROSOFT® WORD® AND MICROSOFT® POWERPOINT®

Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker

Learning Objectives

1. Learn how to paste an image of an Excel chart into a Word document.
2. Learn how to paste a link to an Excel chart into a PowerPoint slide.

Charts that are created in Excel are commonly used in Microsoft Word documents or for presentations that use Microsoft PowerPoint slides. Excel provides options for pasting an image of a chart into either a Word document or a PowerPoint slide. You can also establish a link to your Excel charts so that if you change the data in your Excel file, it is automatically reflected in your Word or PowerPoint files. We will demonstrate both methods in this section.

Pasting a Chart Image into Word

If you wish to follow along, for this exercise you will need two files:

- *The Excel spreadsheet you have been working with in this chapter — **CH4 Charting**.*
- *A Word document data file — CH4 Diversity*

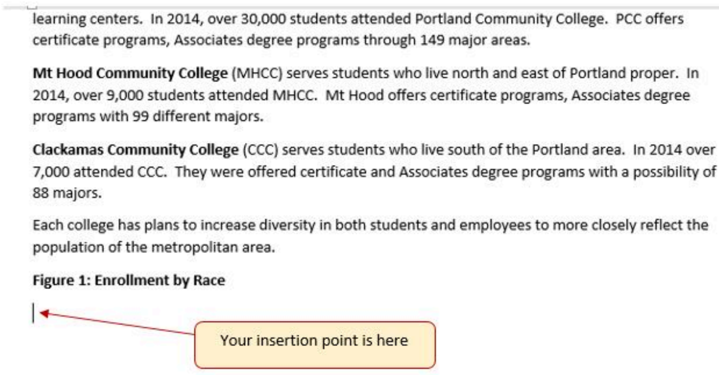
Excel charts can be valuable tools for explaining quantitative data in a written report. Reports that address business plans, public policies, budgets, and so on all involve quantitative data. For this example, we will assume that the Change in Enrollment Statistics Spend Source stacked column chart is being used in a student's written report (see **Figure 2.81**). The following steps demonstrate how to paste an image, or picture, of this chart into a Word document:

1. Open *CH4 Diversity*. Save it as *CH4 Diversity in Enrollment in Community Colleges*
2. Click below the figure heading in the Word document that reads: **Figure 1: Enrollment by Race**. The image of the stacked column chart will be placed below this heading.

3. If needed, open the Excel file you have been working with (*CH4 Charting)*. Activate the **Enrollment by Race** chart in the **Enrollment by Race Chart** sheet.
4. Click the down arrow on the **Copy** button in the **Home** tab of the ribbon. Select **Copy as Picture**
5. Select **OK** — Accepting the Copy Pictures defaults:

    - As shown on Screen
    - Picture

6. Go back to the *CH4 Diversity in Enrollment in Community Colleges* Word document by clicking the file in the taskbar.
7. Confirm that the insertion point is below the Figure 1: Enrollment by Race heading (see Figure 4.42) and click the **Paste** button in the Home tab of the ribbon (or press **Crtl-V**).


learning centers. In 2014, over 30,000 students attended Portland Community College. PCC offers certificate programs, Associates degree programs through 149 major areas.

Mt Hood Community College (MHCC) serves students who live north and east of Portland proper. In 2014, over 9,000 students attended MHCC. Mt Hood offers certificate programs, Associates degree programs with 99 different majors.

Clackamas Community College (CCC) serves students who live south of the Portland area. In 2014 over 7,000 attended CCC. They were offered certificate and Associates degree programs with a possibility of 88 majors.

Each college has plans to increase diversity in both students and employees to more closely reflect the population of the metropolitan area.

Figure 1: Enrollment by Race

Figure 2.81 Paste Picture in Word

Oh no!! The picture is so big that it falls on to the next page. We will need to change its size.

1. Click anywhere on the picture of the chart to activate it.
2. Click the **Format** tab under the **Picture Tools** section of the ribbon (see **Figure 2.82**).
3. Click the down arrow on the **Shape Width** button in the **Size** group of commands. Continue to click the down arrow until the width of the picture is **5.4**." As you reduce the width of the picture, the height is automatically reduced as well. (The height should be ~ **3.92**″)
4. To center the chart on the page, make sure the chart is activated. Then go to the **Home** tab, to the **Paragraph** group, and select **Center.**
5. Save your work.

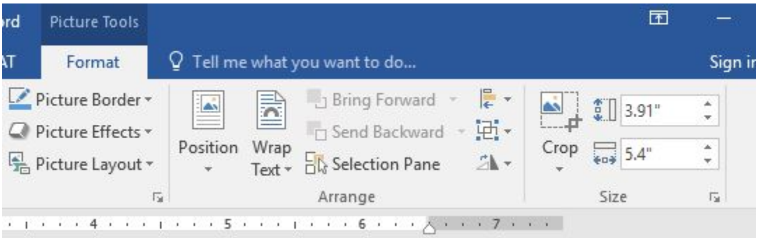


Figure 2.82 Changing the Size of a Picture in Word

Figure 2.83 shows the final appearance of the Enrollment by Race Source chart pasted into a Word document. It is best to use either the Shape Width or Shape Height buttons to reduce the size of the chart. Using either button automatically reduces the height and width of the chart in proper proportion. If you choose to use the sizing handles to resize the chart, holding the SHIFT key while clicking and dragging on a corner sizing handle will also keep the chart in proper proportion.

Diversity in Enrollment in Community Colleges in the Portland Metropolitan Area.

The Portland metropolitan area benefits from a wide array of public and private colleges. By far, most students are enrolled at one of the local community colleges.

Portland Community College (PCC) is the largest, with four full fledges campuses and several smaller learning centers. In 2014, over 30,000 students attended Portland Community College. PCC offers certificate programs, Associates degree programs through 149 major areas.

Mt Hood Community College (MHCC) serves students who live north and east of Portland proper. In 2014, over 9,000 students attended MHCC. Mt Hood offers certificate programs, Associates degree programs with 99 different majors.

Clackamas Community College (CCC) serves students who live south of the Portland area. In 2014 over 7,000 attended CCC. They were offered certificate and Associates degree programs with a possibility of 88 majors.

Each college has plans to increase diversity in both students and employees to more closely reflect the population of the metropolitan area.

Figure 1: Enrollment by Race

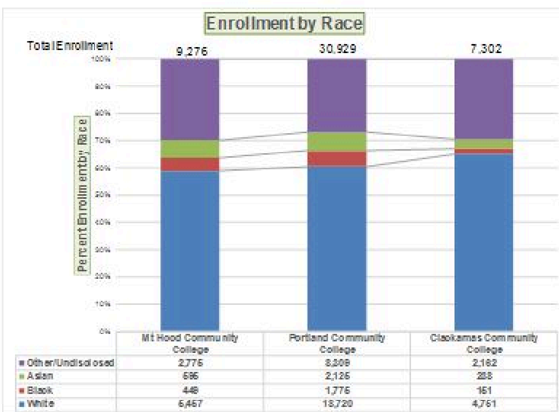


Figure 2.83 Final Appearance of Pasting a Chart Image into Word

Skill Refresher

Pasting a Chart Image into Word

1. Activate an Excel chart and click the Copy button in the Home tab of the ribbon.
2. Click on the location in the Word document where the Excel chart will be pasted.
3. Click the down arrow of the Paste button in the Home tab of the ribbon.
4. Click the Picture option from the drop-down list.
5. Click the Format tab in the Picture Tools section of the ribbon.
6. Resize the picture by clicking the up or down arrow on the Shape Width or Shape Height buttons.

Pasting a Linked Chart Image into PowerPoint

If you wish to follow along, for this exercise you will need two files:

- *The Excel spreadsheet you have been working with in this chapter — **CH4 Charting**.*
- *A PowerPoint data file – CH4 Diversity.*

Microsoft PowerPoint is perhaps the most commonly used tool for delivering live presentations. The charts used in a live presentation are critical for efficiently delivering your ideas to an audience. Similar to written documents, a wide range of presentations may require the explanation of quantitative data. This demonstration includes a PowerPoint slide that could be used in a presentation. We will paste the Enrollment by Race chart into this PowerPoint slide. However, instead of pasting an image, as demonstrated in the Word document, we will establish a link to the Excel file. As a result, if we change the chart in the Excel file, the change will be reflected in the PowerPoint file. The following steps explain how to accomplish this:

1. Open *CH4 Diversity.pptx*. Save it as *CH4 Diversity in Enrollment in Community Colleges*.
2. Navigate to **Slide 6 – Diversity in Enrollment**. This is the slide where you will place the linked chart.
3. If needed, open the Excel file you have been working with (*CH4 Charting)*. Activate the **Enrollment by Race** chart in the **Enrollment by Race Chart** sheet.
4. Click the down arrow on the **Copy** button in the **Home** tab of the ribbon. Select **Copy** (not Copy as Picture.)

5. Go back to the *CH4 Diversity in Enrollment in Community Colleges* presenation by clicking the file in the taskbar.
6. Make sure you are still on Slide 6 – Diversity in Enrollment. Click on the **outside edge** of the empty prompt box on the right.
7. Click the down arrow below the **Paste** button in the **Home** tab of the ribbon in the PowerPoint file.
8. Hover over each of the Paste Options until you find **Keep Source Formatting & Link Data** (see Figure 2.84). Select this option. This pastes an image of the Excel chart into the PowerPoint slide. In addition, a link is created so that any changes made to the chart (in Excel) appear on the PowerPoint slide.

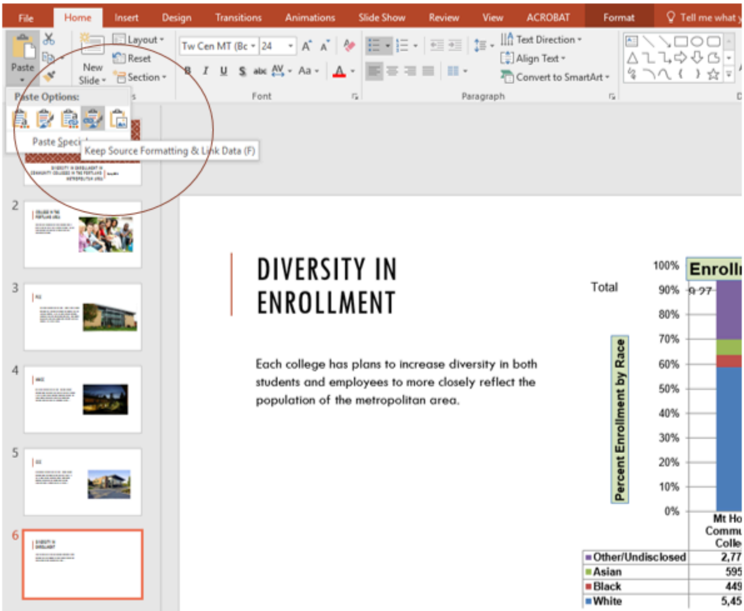


Figure 2.84 Creating a Link to an Excel Chart in PowerPoint

Next we need to make some changes to clean up the chart a bit. First, we are going to apply a different chart style.

1. Click anywhere in the plot area of the column chart pasted into the PowerPoint slide. You will see the same **Excel Chart Tools** tabs added to the ribbon (see **Figure 2.85**).
2. On the **Design** tab, select **Style 8** in the **Chart Style** group.

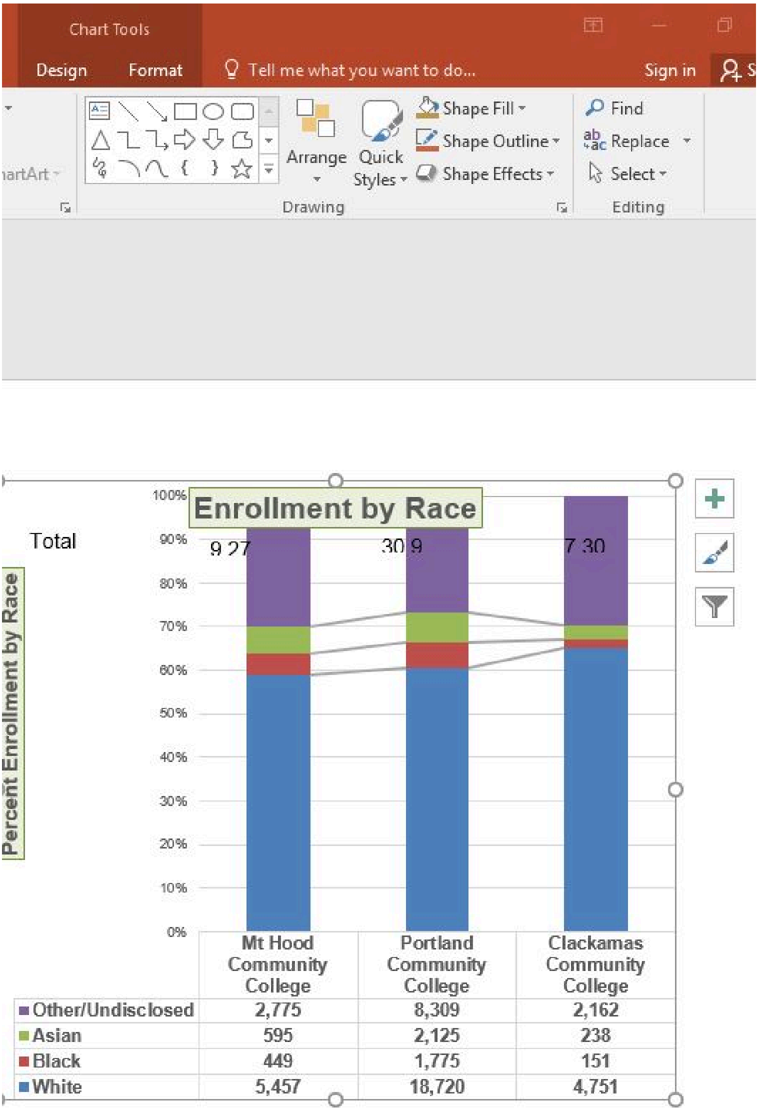


Figure 2.85 Modifying and Excel Chart Pasted into a PowerPoint Slide

Paste linking this chart caused trouble with the text boxes we added, so next we are going to delete them.

1. Select each text box by clicking on the outside edge of the text box with the four-headed arrow. Press the delete key on your keyboard. Be sure that the insertion point is NOT blinking inside the text box. If it is, you will be editing the contents of the text box instead of deleting the actual text box.

The benefit of adding this chart to the presentation as a link is that it will automatically update when you change the data in the linked spreadsheet file.

1. Return to your *CH4 Charting* Excel file.
2. Select the **Enrollment Statistics** worksheet (the one with the Enrollment data.) Change the value in cell D3 to **1000**. You have just changed the number of white students at Clackamas Community College to **1000**. This isn't true, but you want to change the data enough to see the effect in the charts.
3. Select the **Enrollment by Race Chart** worksheet. Notice how the chart has changed.
4. Return to the Diversity in Enrollment in Community Colleges PowerPoint file by clicking the file in the taskbar.
5. On Slide6, you should see the updated chart (see **Figure**

2.86).

6. If the chart has not changed; be sure that your chart is selected, click the **Design** tab in the **Chart Tools** section of the ribbon. Click the **Refresh Data** button. The change made in the Excel workbook is now reflected on the PowerPoint slide.
7. If that still doesn't work, you may have created a "normal" link — instead of a **Paste Link**. Delete the chart and follow the steps again. Start from the beginning of this section.
8. Save your work. You will submit both the Word and PowerPoint files, along with the Excel file, at the end of the next section.

Figure 4.47 shows the appearance of the column chart after the change was made in the Enrollment Statistics worksheet in the Excel file. Note that the Data Chart at the bottom reflects the new number, too. The change that was made in the Excel file will appear in the PowerPoint file after clicking the Refresh Data button.

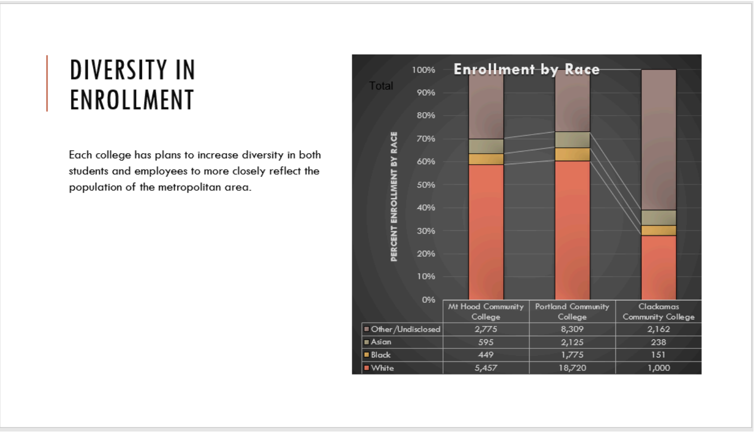


Figure 2.86 Styled and Updated Chart.

Integrity Check

Refreshing Linked Charts in PowerPoint and Word

When creating a link to a chart in Word or PowerPoint, you must refresh the data if you make any changes in the Excel workbook. This is especially true if you make changes in the Excel file prior to opening the Word or PowerPoint file that contains a link to a chart. To refresh the chart, make sure it is activated, then click the

Refresh Data button in the Design tab of the ribbon. Forgetting this step can result in old or erroneous data being displayed on the chart.

Integrity Check

Severed Link?

When creating a link to an Excel chart in Word or PowerPoint, you must keep the Excel workbook in its original location on your computer or network. If you move or delete the Excel workbook, you will get an error message when you try to update the link in your Word or PowerPoint file. You will also get an error if the Excel workbook is saved on a network drive that your computer cannot access. These errors occur because the link to the Excel workbook has been severed. Therefore, if you know in advance that you will be using a USB drive to pull up your documents or presentation, move the Excel workbook to your USB drive before you establish the link in your Word or PowerPoint file.

Skill Refresher:

Pasting a Linked Chart Image into PowerPoint

1. Activate an Excel chart and click the Copy button in the Home tab of the ribbon.
2. Click in the PowerPoint slide where the Excel chart will be pasted.
3. Click the down arrow of the Paste button in the Home tab of the ribbon.
4. Click the Keep Source Formatting & Link Data option from the drop-down list.
5. Click the Refresh Data button in the Design tab of the ribbon to ensure any changes in the Excel file are reflected in the chart.

Key Takeaways

- When pasting an image of an Excel chart into a Word document or PowerPoint file, use the **Picture** option from the Paste drop-down list of options – if you want the image to act as an image. You will not be able to make any changes to the content of the picture.
- When creating a link to a chart in Word or PowerPoint, you may need to refresh the data if you make any changes in the originating spreadsheet. You should not use the **Picture** option.

Attribution

Media Attributions

- Figure 4-42
- Figure 4-43
- Figure 4-44
- fig45
- Figure 4-46
- fig47

TABLE BASICS

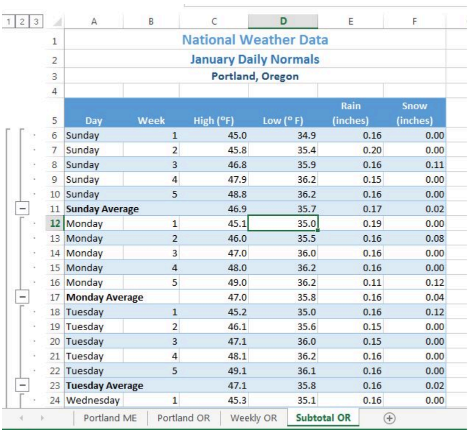
Noreen Brown, Barbara Lave, Julie Romey, Mary Schatz, and Diane Shingledecker

Learning Objectives

1. Understand table structure.
2. Plan, create, and edit a table.
3. Freeze rows and columns.
4. Sort data in a table.

This section reviews the fundamental skills for setting up and maintaining an Excel table. The objective used for this chapter is the construction of a multi-sheet file to keep track of two cities' national weather data for the month of January. Organizing, maintaining, and reporting data are essentials skills for employees in most industries.

Figure 2.91 shows the completed workbook that will be demonstrated in this chapter. Notice that this workbook contains three worksheets. The first worksheet lists average weather for January in Portland, Maine. The second sheet lists average weather data for January in a very different climate – Portland, Oregon. The third sheet adds a weekly column to the Portland, Oregon data so that it can be subtotaled by week.



National Weather Data
January Daily Normals
Portland, Oregon

Day	Week	High (°F)	Low (° F)	Rain (inches)	Snow (inches)
Sunday	1	45.0	34.9	0.16	0.00
Sunday	2	45.8	35.4	0.20	0.00
Sunday	3	46.8	35.9	0.16	0.11
Sunday	4	47.9	36.2	0.15	0.00
Sunday	5	48.8	36.2	0.16	0.00
Sunday Average		46.9	35.7	0.17	0.02
Monday	1	45.1	35.0	0.19	0.00
Monday	2	46.0	35.5	0.16	0.08
Monday	3	47.0	36.0	0.16	0.00
Monday	4	48.0	36.2	0.16	0.00
Monday	5	49.0	36.2	0.11	0.12
Monday Average		47.0	35.8	0.16	0.04
Tuesday	1	45.2	35.0	0.16	0.12
Tuesday	2	46.1	35.6	0.15	0.00
Tuesday	3	47.1	36.0	0.15	0.00
Tuesday	4	48.1	36.2	0.16	0.00
Tuesday	5	49.1	36.1	0.16	0.00
Tuesday Average		47.1	35.8	0.16	0.02
Wednesday	1	45.3	35.1	0.16	0.00

Figure 2.91 Completed National Weather Workbook

Creating a Table

If you wish to follow along, download Data file: [CH5 Data]

When data is presented in long lists or columns, it helps if the table is set up well. Here are some rules of data-entry etiquette to follow when creating a table from scratch:

1. Whenever you can, organize your information using adjacent (neighboring) columns and rows.
2. Start the table in the upper-left corner of the worksheet and work your way down the sheet.

3. Don't skip columns and rows just to "space out" the information. (To place white space between information in adjacent columns and rows, you can widen columns, heighten rows, and change the alignment.)
4. Reserve a single column at the left edge of the table for the table's row headings or identifying information.
5. Reserve a single row at the top of the table for the table's column headings.
6. If your table requires a title, put the title in the row(s) above the column headings.

Following these rules will help insure that the sorts, filters, totals, and subtotals you apply to your table with give you the desired results.

With these rules in mind, we will begin working on the **Portland ME** worksheet in the **National Weather workbook**. Notice that the data is in adjacent columns and rows. The upper-left corner of the table is in A5 and the titles are above the column headings in Row 5. Since the set-up of our data looks good, we are ready to turn our data range into an Excel table:

1. Open data file *CH5 Data* and save a file to your computer as **CH5 National Weather**.
2. Click on A5 in the **Portland ME** sheet.
3. Click the Table button in the Insert tab of the Ribbon. **Figure 2.92** will appear on your screen.

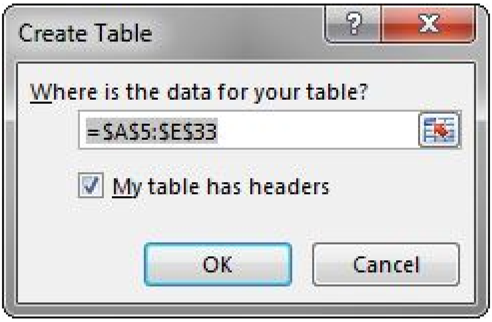
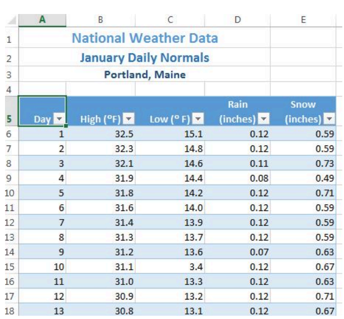


Figure 2.92 Create Table

1. Make sure “My table has headers” is checked. Click OK.
2. Click in A5 again.
3. Adjust your columns widths so that you can see the complete headings in row 5 with the filter arrows showing. The filter arrows are the down-arrow buttons that will appear in row 5 when you create your table. We will learn how to use these to sort and filter later in this chapter.

After this, your spreadsheet will look like **Figure 2.93**.



	A	B	C	D	E
1		National Weather Data			
2		January Daily Normals			
3		Portland, Maine			
4					
5	Day	High (ºF)	Low (º F)	Rain (inches)	Snow (inches)
6	1	32.5	15.1	0.12	0.59
7	2	32.3	14.8	0.12	0.59
8	3	32.1	14.6	0.11	0.73
9	4	31.9	14.4	0.08	0.49
10	5	31.8	14.2	0.12	0.71
11	6	31.6	14.0	0.12	0.59
12	7	31.4	13.9	0.12	0.59
13	8	31.3	13.7	0.12	0.59
14	9	31.2	13.6	0.07	0.63
15	10	31.1	3.4	0.12	0.67
16	11	31.0	13.3	0.12	0.63
17	12	30.9	13.2	0.12	0.71
18	13	30.8	13.1	0.12	0.67

Figure 2.93 Weather Table

Notice that a new ribbon tab, Table Tools Design, appears when you click inside your table. This ribbon tab allows you to edit, style, and add functionality to your table.

Let's try these steps again in the following steps:

1. Click on the **Portland OR** sheet and click in cell A5.
2. Click the Table button in the Insert tab of the Ribbon.
3. Make sure "My table has headers" is checked. Click OK.
4. Click in A5 again.
5. Adjust your columns widths so that you can see the complete headings in row 5 with the filter arrows showing.

Skill Refresher

Create a Table

1. Click on the top left cell in your data.
2. Click the Table button in the Insert tab of the Ribbon.
3. Make sure "My table has headers" is checked. Click OK.
4. Click on the top left cell again.

5. Adjust your columns widths so that you can see the complete headings with the filter arrows showing.

Formatting Tables

There are many ways to format an Excel table. There are preset colored Table Styles with Light, Medium, and Dark colors. There are also a variety of Table Style Options listed in **Table 5.1**.

Table 2.91 Table Style Options

Table Style	Description
Header Row	Top row of the table that includes column headings
Total Row	Row added to the bottom that applies column summary calculations
First Column	Formatting added to the left-most column in the table
Last Column	Formatting added to the right-most column in the table
Banded Rows	Alternating rows of color added to make it easier to see rows of data
Banded Columns	Alternating columns of color added to make it easier to see columns of data
Filter Button	Button that appear at the top of each column that lists options for sorting and filtering

We'll add some formatting to both of our Portland weather tables in the following steps:

1. Click on the **Portland ME** sheet in your file.
2. In the Table Tools Design tab, in the Table Styles group, click the More button. 

A gallery of table styles will appear as in **Figure 2.94**.

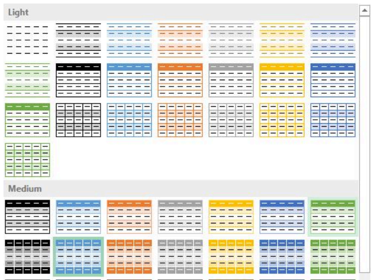


Figure 2.94 Table Styles

3. In the Table Styles gallery, in the Medium Section, click Table Style Medium 7.

4. In the Table Style Options group in the Ribbon, click Banded Rows.

The alternating colored rows will disappear. The data in the table is now more difficult to read.

5. Try out some of the other options in the Table Style Options group. Once you're finished, check just Header Row, Banded Rows, and Filter Button as in **Figure 2.95** below.

Figure 2.95 Ribbon Table Style Options

Adding Data to Tables

Over time, you will need to add new data to an Excel table. You will add the data to the table in a blank row. The easiest way to do this is to enter the data in the first blank row below the last row in the table. You can then rearrange the data in the table by sorting it. If you need to add data in a specific place in the middle of a table, you can insert a blank row in the middle and add your data there.

We need to add the last three days of the months to both our Portland, Maine and Portland, Oregon tables. The following steps will walk you through doing this.

1. Click on the **Portland ME** worksheet.
2. Click on A34 (the left-most cell below the last row in the table).
3. Enter the following data:

Table 2.92 Portland, Maine data

Day	High (°F)	Low (°F)	Rain (inches)	Snow (inches)
29	31.4	13.3	0.12	0.59
30	31.6	3.4	0.08	0.47
31	31.7	13.5	0.12	0.63

Notice that the banded row formatting continues as additional rows are added to the tables.

1. Click on the **Portland OR** worksheet.
2. Click on A34 (the left-most cell below the last row in the table).
3. Enter the following data:

Table 2.93 Portland, Oregon data

Day	High (°F)	Low (°F)	Rain (inches)	Snow (inches)
29	48.8	36.2	0.16	0
30	49.0	36.2	0.11	0.32
31	49.1	36.1	0.16	0

Finding and Editing Data

It is inevitable that you will find data errors in your table and need to correct them. While you can visually scan through a table to find your errors, this can be a tedious and tiresome process. Excel can help with this through the Find command. When you use Find, the best practice is to start at the top of the table to ensure that all your data is included in the search.

We know that a temperature of 3.4 degrees (brrr!) was entered erroneously in the **Portland Maine sheet**. It should have been 13.4. To fix this error, complete the following steps.

1. Click on the **Portland ME** sheet.
2. Press the CTRL+HOME keys together to go to the top of the sheet (A1).
3. In the Home tab of the ribbon, click on Find & Select in the Editing Group and then click Find.
4. In the Find box, type **3.4**, and then click Find Next.

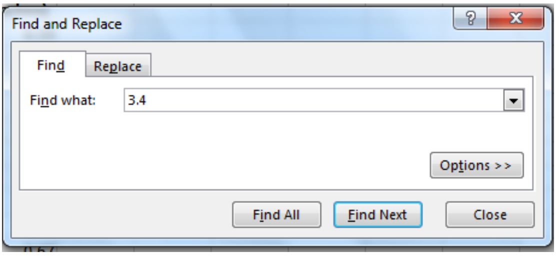


Figure 5.6 Find and Replace

5. Click the Close button.
6. Replace 3.4 in the Low column for Day 10 with **13.4.**
7. Now switch to the **Portland Oregon** sheet and find the Snow error of .32. Change it to **0.12**. You should find the error in Day 3.

Skill Refresher

Finding and Replacing Data

1. In the Home tab of the ribbon, click on Find & Select in the Editing Group and then click Find.
2. In the Find box, type what you want to find, and then click Find Next.
3. Continuing click Find Next until you find.what you are looking for.
4. Click Close and edit your data.

Freeze Rows and Columns

When you freeze panes, Microsoft Excel keeps specific rows or columns visible in your table when you scroll through it on your screen. For example, if the first row in your spreadsheet contains labels, you might freeze that row to make sure that the column labels remain visible as you scroll down in your spreadsheet. When we scroll through our weather data, it would be nice to keep our column headings visible on the screen.

To freeze your headings:

1. Click in A6, the left-most cell below the headings row.
2. Click the View tab in the ribbon.
3. Select Freeze Panes and then Freeze Panes again.
4. Scroll up and down the sheet and notice that the headings are always displayed at the top of the table.

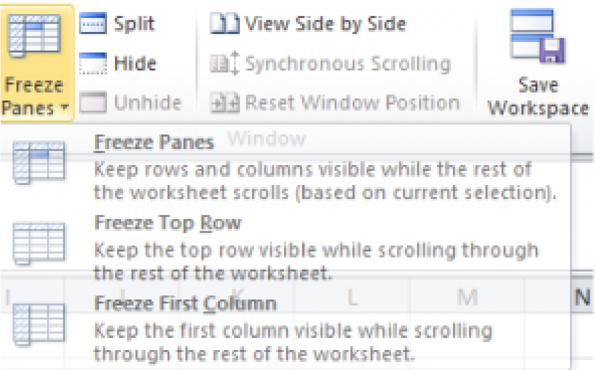


Figure 2.97 Freeze Pane

To unfreeze your headings:

1. Click on the View tab in the ribbon.
2. Select Unfreeze Panes.

Simple Sort

Content in a table can be sorted alphabetically, numerically, and in many other ways. Sorting helps organize data by one or more columns in your table. **Table 5.4** describes the different sort orders available for each column of data.

Table 2.94 Sort Options

Sort Order	Text	**Numbers**	**Dates**
Ascending	Alphabetical (A-Z)	Smallest to Largest Lowest to Highest	Chronological (oldest to newest)
Descending	Reverse Alphabetical (Z-A)	Largest to Smallest Highest to Lowest	Reverse Chronological (newest to oldest)

Let's say we want to know what the snowiest day was in January in Portland, Maine; so we want to sort the Snow column in Descending order so that the snowiest day ends up at the top of the table.

1. Click on the filter Click arrow to the right of the header Snow (inches) in the **Portland ME** worksheet.
2. Click on the choice Click ZA↓ Sort Largest to Smallest. See **Figure 2.98** below.

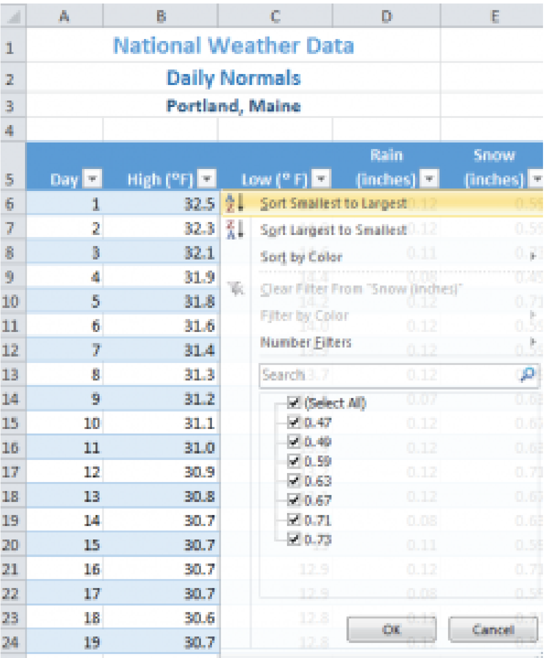


Figure 2.98 Sort by One Column

If you did this correctly, you'll see that the snowiest day at the top of the list in January 3rd (in row 6) with 0.73 inches of snow! Notice the filter arrow changes in the snow column to a downward pointing arrow to indicate you sorted that column in descending order (largest to smallest).

National Weather Data

January Daily Normals

Portland, Maine

Day	High (°F)	Low (° F)	Rain (inches)	Snow (inches)
3	32.1	14.6	0.11	0.73
5	31.8	14.2	0.12	0.71
12	30.9	13.2	0.12	0.71
16	30.7	12.9	0.12	0.71
18	30.6	12.8	0.12	0.71
22	30.8	12.8	0.08	0.71
10	31.1	13.4	0.12	0.67

Figure 2.99 Snowiest Days in Maine

3. Now switch to the **Portland Oregon** sheet and repeat these sort steps to find the snowiest day in Oregon. Check your answers with **Figure 2.910**.

National Weather Data

January Daily Normals

Portland, Oregon

Day	High (°F)	Low (° F)	Rain (inches)	Snow (inches)
30	49.0	36.2	0.11	0.32
3	45.2	35.0	0.16	0.12
15	46.8	35.9	0.16	0.11
9	46.0	35.5	0.16	0.08
25	48.3	36.2	0.12	0.08
1	45.0	34.9	0.16	0.00
2	45.1	35.0	0.19	0.00

Figure 2.910 Snowiest Days in Oregon

Skill Refresher

Sort a Column

1. Click on the filter Click arrow to the right of the header in the column you want to sort.
2. Click on the choice AZ! or ZA↓ to sort your data by that column.

Key Takeaways

- Tables are made up of adjacent rows and columns of data with a single row of column headings at the top.
- You can create a table by clicking in the top left-most cell in your data and selecting Table in the Insert tab of the ribbon.
- There are a gallery of styles, as well as, style options to choose from to format a table.
- When you need to add data, it is best to add

it one row below the bottom of the table. You can then sort to reorganize your data.

- Freezing heading keeps your column headings displayed while you scroll through your table data.
- You can use the filter arrows in the table headings to sort by a single column.

Attribution

Media Attributions

- figure-5-5-table-style-options
- figure-5-7-freeze-panes
- figure-5-8-sort
- figure-5-9-snowiest-maine
- figure-5-10-snowiest-oregon

CHAPTER III

CHAPTER 3: MATH OF FINANCE

WHY IT MATTERS: FINANCE

Why study interest formulas?

You’re in the market for a new refrigerator, but don’t have a lot of cash on hand to make the purchase. A flyer from an appliance rent-to-own store arrives in the mail one day, containing a very tempting offer: the refrigerator of your dreams for only $17.99 per week!

The thought of paying $17.99 a week seems reasonable given your current budget, but you hesitate when you read the fine print. The rent-to-own contract specifies that payments must be made for two full years. That's 104 weeks at $17.99 per week!

At the local big box store, the same refrigerator is listed at only $1299, including all taxes and fees. When you tell your brother about the two deals, he offers to help you buy the refrigerator from the big box store at the lower price of $1299. However he will charge you 20% interest on the full price and wants you to pay off the balance within 12 months. You like the lower price, but 20% seems like a pretty high percentage to pay out to your brother.

Then you discover a third option. The big box store offers a store credit line at 15% APR. After reading the fine print, you learn that the credit line works just like a loan. The interest will be compounded each month, and there will be a fixed monthly payment for a total of 36 months. You wonder how much interest will accumulate on the $1299 ticket price of the refrigerator.

Which offer is better? Renting-to-own for two years, buying it on a 20% loan from your brother, or using the store's line of credit at 15% compounding interest? Better think quickly: your ice cream is melting!

In order to make an informed decision, you will need to know the total cost for all three scenarios. The rent-to-own situation is the easiest to calculate because all of the fees and interest have been figured into the monthly payment already. Simply multiply the number of weeks in two full years by the weekly payment.

$$104 \times 17.99 = 1870.96$$

The other two scenarios involve interest formulas. We will revisit this scenario to see which offer is the best deal after taking a look at the other options.

In this module, you will learn two ways to calculate interest; simple and compound. Understanding interest rates will help you become a more informed consumer, potentially saving you a lot of money on big purchases such as appliances, cars and even your home.

Attributions

SIMPLE AND COMPOUND INTEREST

Learning Outcomes

- Calculate one-time simple interest, and simple interest over time
- Determine APY given an interest scenario
- Calculate compound interest

We have to work with money every day. While balancing your checkbook or calculating your monthly expenditures on espresso requires only arithmetic, when we start saving, planning for retirement, or need a loan, we need more mathematics.

Simple Interest

Discussing interest starts with the **principal**, or amount your account starts with. This could be a starting investment, or the starting amount of a loan. Interest, in its most simple form, is calculated as a percent of the principal. For example, if you borrowed \$100 from a friend and agree to repay it with 5% interest, then the amount of interest you would pay would just be 5% of 100: \$100(0.05) = \$5. The total amount you would repay would be \$105, the original principal plus the interest.

Simple One-time Interest

$$I = P_0 r$$

(1) $$A = P_0 + I = P_0 + P_0 r = P_0(1 + r)$$

- *I* is the interest
- *A* is the end amount: principal plus interest
- (2) P_0

  is the principal (starting amount)
- *r* is the interest rate (in decimal form. Example: 5% = 0.05)

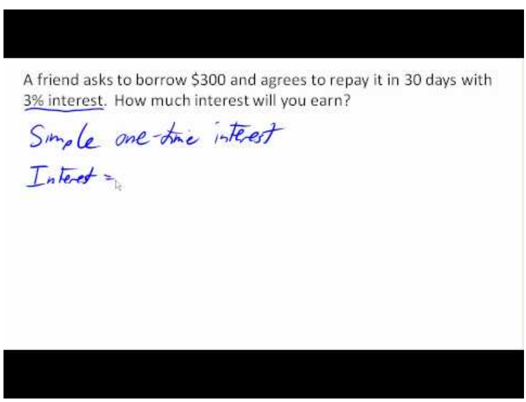
Examples

A friend asks to borrow $300 and agrees to repay it in 30 days with 3% interest. How much interest will you earn?

Solution:

(3) P_0 = \$300	the principal
$r = 0.03$	3% rate
$I = \$300(0.03) = \$9.$	You will earn \$9 interest.

The following video works through this example in detail.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=133

One-time simple interest is only common for extremely short-term loans. For longer term loans, it is common for interest to be paid on a daily, monthly, quarterly, or annual basis. In that case, interest would be earned regularly.

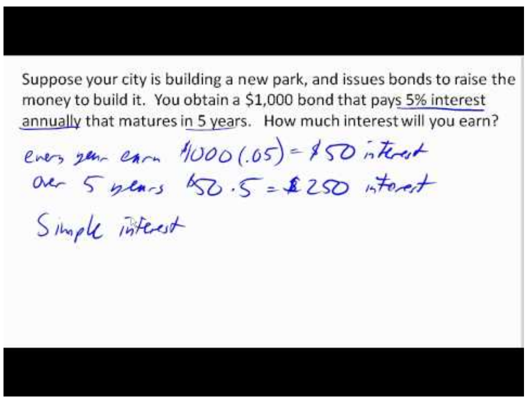
For example, bonds are essentially a loan made to the bond issuer (a company or government) by you, the bond holder. In return for the loan, the issuer agrees to pay interest, often annually. Bonds have a maturity date, at which time the issuer pays back the original bond value.

Exercises

Suppose your city is building a new park, and issues bonds to raise the money to build it. You obtain a $1,000 bond that pays 5% interest annually that matures in 5 years. How much interest will you earn?
[reveal-answer q=”14596"]Show Solution[/reveal-answer]
[hidden-answer a=”14596"]Each year, you would earn 5% interest: $1000(0.05) = $50 in interest. So over the course of five years, you would earn a total

of $250 in interest. When the bond matures, you would receive back the $1,000 you originally paid, leaving you with a total of $1,250.[/hidden-answer]

Further explanation about solving this example can be seen here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=133

We can generalize this idea of simple interest over time.

Simple Interest over Time

(4)

$$I = P_0 rt$$
$$A = P_0 + I = P_0 + P_0 rt = P_0(1 + rt)$$

- *I* is the interest
- *A* is the end amount: principal plus interest
- (5) P_0

  is the principal (starting amount)
- *r* is the interest rate in decimal form
- *t* is time

The units of measurement (years, months, etc.) for the time should match the time period for the interest rate.

APR – Annual Percentage Rate

Interest rates are usually given as an **annual percentage rate (APR)** – the total interest that will be paid in the year. If the interest is paid in smaller time increments, the APR will be divided up.

For example, a 6% APR paid monthly would be divided into twelve 0.5% payments.
$6 \div 12 = 0.5$

A 4% annual rate paid quarterly would be divided into four 1% payments.
$4 \div 4 = 1$

Example

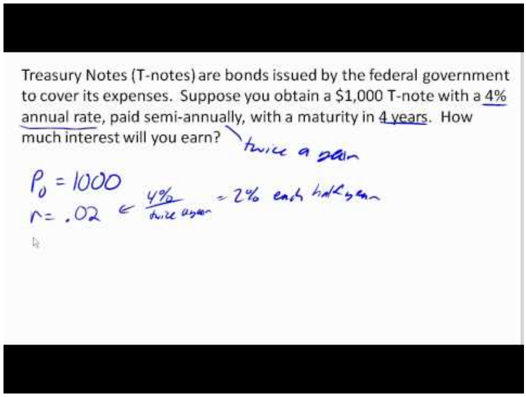
Treasury Notes (T-notes) are bonds issued by the federal government to cover its expenses. Suppose you obtain a $1,000 T-note with a 4% annual rate, paid semi-annually, with a maturity in 4 years. How much interest will you earn?

Solution:

Since interest is being paid semi-annually (twice a year), the 4% interest will be divided into two 2% payments.

(6) P_0 = \$1000	the principal
$r = 0.02$	2% rate per half-year
$t = 8$	4 years = 8 half-years
$I = \$1000(0.02)(8) = \$160.$	You will earn \$160 interest total over the four years.

This video explains the solution.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=133

Try It

[Click here to try this problem.]()

Try It

A loan company charges $30 interest for a one month loan of $500. Find the annual interest rate they are charging.

Solution:

I = $30 of interest
P_0 = $500 principal
r = unknown
t = 1 month

Using $I = P_0 rt$, we get $30 = 500r1$. Solving, we get r = 0.06, or 6%. Since the time was monthly, this is the monthly interest. The annual rate would be 12 times this: 72% interest.

Try It

Click here to try this problem.

Compound Interest

With simple interest, we were assuming that we pocketed the interest when we received it. In a standard bank account, any interest we earn is automatically added to our balance, and we earn interest on that interest in future years. This reinvestment of interest is called **compounding**.

Suppose that we deposit \$1000 in a bank account offering 3% interest, compounded monthly. How will our money grow?

The 3% interest is an annual percentage rate (APR) – the total interest to be paid during the year. Since interest is being paid monthly, each month, we will earn 3% ÷ 12 = 0.25% per month.

In the first month,

- $P_0 = \$1000$
- $r = 0.0025$ (0.25%)
- $I = \$1000\ (0.0025) = \$2.50$
- $A = \$1000 + \$2.50 = \$1002.50$

In the first month, we will earn $2.50 in interest, raising our account balance to $1002.50.

In the second month,

- $P_0 = \$1002.50$
- $I = \$1002.50\ (0.0025) = \$2.51$ (rounded)
- $A = \$1002.50 + \$2.51 = \$1005.01$

Notice that in the second month we earned more interest than we did in the first month. This is because we earned interest not only on the original $1000 we deposited, but we also earned interest on the $2.50 of interest we earned the first month. This is the key advantage that **compounding** interest gives us.

Calculating out a few more months gives the following:

Month	Starting balance	Interest earned	Ending Balance
1	1000.00	2.50	1002.50
2	1002.50	2.51	1005.01
3	1005.01	2.51	1007.52
4	1007.52	2.52	1010.04
5	1010.04	2.53	1012.57
6	1012.57	2.53	1015.10
7	1015.10	2.54	1017.64
8	1017.64	2.54	1020.18
9	1020.18	2.55	1022.73
10	1022.73	2.56	1025.29
11	1025.29	2.56	1027.85
12	1027.85	2.57	1030.42

We want to simplify the process for calculating compounding, because creating a table like the one above is time consuming. Luckily, math is good at giving you ways to take shortcuts. To find an equation to represent this, if P_m represents the amount of money after m months, then we could write the recursive equation:

$P_0 = \$1000$

$P_m = (1+0.0025)P_{m-1}$

You probably recognize this as the recursive form of exponential growth. If not, we go through the steps to build an explicit equation for the growth in the next example.

Example

Build an explicit equation for the growth of \$1000 deposited in a bank account offering 3% interest, compounded monthly.

Solution:

- P_0 = \$1000
- $P_1 = 1.0025P_0$ = 1.0025 (1000)
- $P_2 = 1.0025P_1$ = 1.0025 (1.0025 (1000)) = 1.0025 2(1000)
- $P_3 = 1.0025P_2$ = 1.0025 (1.00252(1000)) = 1.00253(1000)
- $P_4 = 1.0025P_3$ = 1.0025 (1.00253(1000)) = 1.00254(1000)

Observing a pattern, we could conclude

- $P_m = (1.0025)^m$(\$1000)

Notice that the \$1000 in the equation was P_0, the starting amount. We found 1.0025 by adding one to the growth rate divided by 12, since we were compounding 12 times per year.

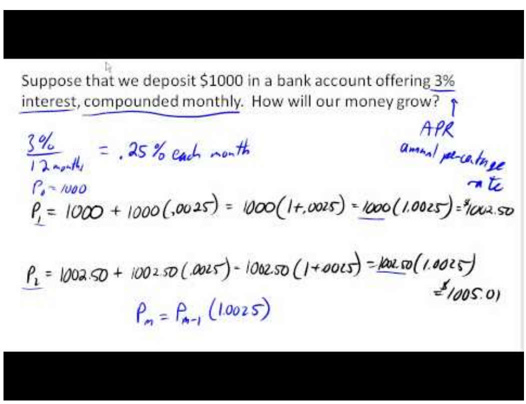
Generalizing our result, we could write

$$P_m = P_0\left(1+\frac{r}{k}\right)^m$$

In this formula:

- *m* is the number of compounding periods (months in our example)
- *r* is the annual interest rate
- *k* is the number of compounds per year.

View this video for a walkthrough of the concept of compound interest.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=133

While this formula works fine, it is more common to use a formula that involves the number of years, rather than the number of compounding periods. If N is the number of years, then $m = N\ k$. Making this change gives us the standard formula for compound interest.

Compound Interest

$$P_N = P_0\left(1 + \frac{r}{k}\right)^{Nk}$$

- P_N is the balance in the account after N years.
- P_0 is the starting balance of the account (also called initial deposit, or principal)
- r is the annual interest rate in decimal form
- k is the number of compounding periods in one year
    - If the compounding is done annually (once a year), $k = 1$.
    - If the compounding is done quarterly, $k = 4$.
    - If the compounding is done monthly, $k = 12$.
    - If the compounding is done daily, $k =$

365.

The most important thing to remember about using this formula is that it assumes that we put money in the account once and let it sit there earning interest.

In the next example, we show how to use the compound interest formula to find the balance on a certificate of deposit after 20 years.

Example

A certificate of deposit (CD) is a savings instrument that many banks offer. It usually gives a higher interest rate, but you cannot access your investment for a specified length of time. Suppose you deposit \$3000 in a CD paying 6% interest, compounded monthly. How much will you have in the account after 20 years?

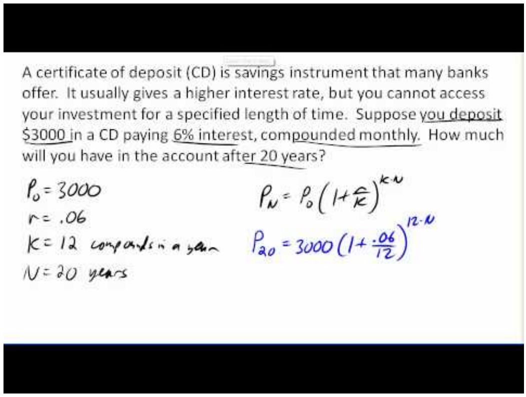
Solution:

In this example,

P_0 = \$3000	the initial deposit
r = 0.06	6% annual rate
k = 12	12 months in 1 year
N = 20	since we're looking for how much we'll have after 20 years

So $P_{20} = 3000\left(1 + \frac{0.06}{12}\right)^{20\times 12} = \$9930.61$ (round your answer to the nearest penny)

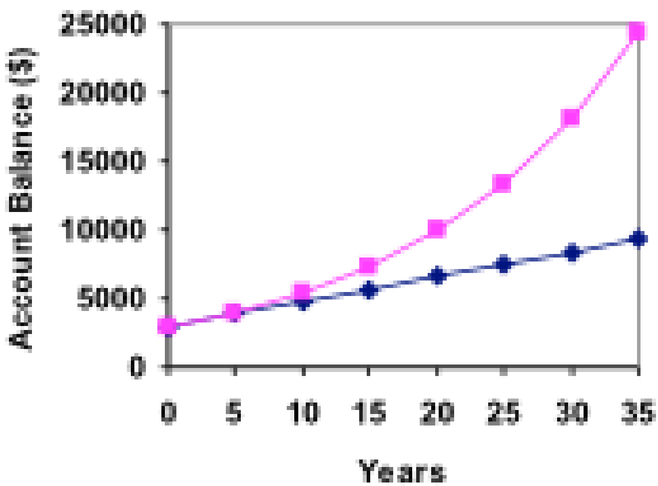
A video walkthrough of this example problem is available below.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=133

Let us compare the amount of money earned from compounding against the amount you would earn from simple interest

Years	Simple Interest ($15 per month)	6% compounded monthly = 0.5% each month.
5	$3900	$4046.55
10	$4800	$5458.19
15	$5700	$7362.28
20	$6600	$9930.61
25	$7500	$13394.91
30	$8400	$18067.73
35	$9300	$24370.65



As you can see, over a long period of time, compounding makes a large difference in the account balance. You may recognize this as the difference between linear growth and exponential growth.

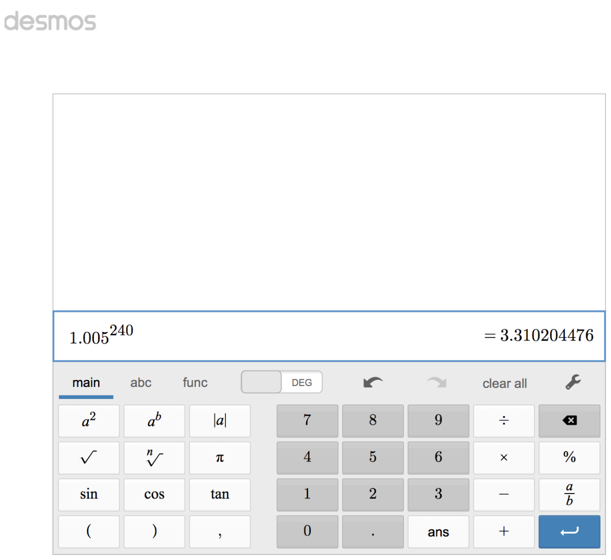
Try It

Click here to try this problem.

Evaluating exponents on the Desmos calculator

When we need to calculate something like 5^3 it is easy enough to just multiply $5 \cdot 5 \cdot 5 = 125$. But when we need to calculate something like 1.005^{240}, it would be very tedious to calculate this by multiplying 1.005 by itself 240 times! So to make things easier, we can harness the power of our scientific calculators. In this class, we are using the Desmos calculator. If you just want to square a number, the key is a^2. If you want to raise a number to another power, you use the key a^b on the main menu.

To evaluate 1.005^{240} we'd type 1.005 a^b 240 . Try it out – you should get the answer in the figure below:



Most scientific calculators have a button for exponents. If you are not using the Desmos calculator, it is typically either labeled like:

^ , y^x , or x^y .

Example

You know that you will need $40,000 for your child's education in 18 years. If your account earns 4% compounded quarterly, how much would you need to deposit now to reach your goal?

Solution:

In this example, we're looking for P_0.

$r = 0.04$	4%
$k = 4$	4 quarters in 1 year
$N = 18$	Since we know the balance in 18 years
$P_{18} = \$40,000$	The amount we have in 18 years

In this case, we're going to have to set up the equation, and solve for P_0.

$$40000 = P_0\left(1+\frac{0.04}{4}\right)^{18\times 4}$$

$$40000 = P_0(2.0471)$$

$$P_0 = \frac{40000}{2.0471} = \$19539.84 \tag{7}$$

So you would need to deposit $19,539.84 now to have $40,000 in 18 years.

Try It

Click here to try this problem.

Rounding

If you are not inputting your entire formula into Desmos and choose to do it section by section, It is important to be very careful about rounding when calculating things with exponents. In general, you want to keep as many decimals during calculations as you can. Be sure to **keep at least 3 significant digits** (numbers after any leading zeros). Rounding 0.00012345 to 0.000123 will usually give you a "close enough" answer, but keeping more digits is always better.

Example

To see why not over-rounding is so important, if you choose not to enter your formula all at once into Desmos, suppose you were investing $1000 at 5% interest compounded monthly for 30 years.

P_0 = \$1000	the initial deposit
r = 0.05	5%
k = 12	12 months in 1 year
N = 30	since we're looking for the amount after 30 years

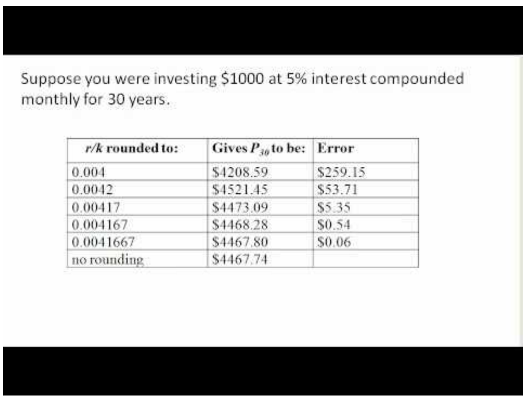
If we first compute r/k, we find 0.05/12 = 0.00416666666667

Here is the effect of rounding this to different values:

r/k **rounded to:**	**Gives** ***P30*** **to be:**	**Error**
0.004	\$4208.59	\$259.15
0.0042	\$4521.45	\$53.71
0.00417	\$4473.09	\$5.35
0.004167	\$4468.28	\$0.54
0.0041667	\$4467.80	\$0.06
no rounding	\$4467.74	

If you're working in a bank, of course you wouldn't round at all. For our purposes, the answer we got by rounding to 0.00417, three significant digits, is close enough – \$5 off of \$4500 isn't too bad. Certainly keeping that fourth decimal place wouldn't have hurt.

View the following for a demonstration of this example.

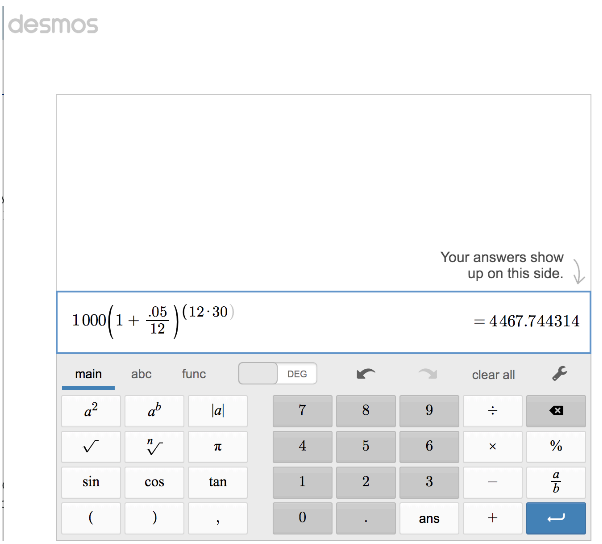


A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=133

Using your Desmos calculator

In many cases, you can avoid rounding completely by how you enter things in your calculator. For example, in the example above, we needed to calculate $P_{30} = 1000\left(1 + \frac{0.05}{12}\right)^{12\times30}$

We can quickly calculate this on the Desmos Calculator by putting in the formula all at once:



To enter this into the calculator, type in the following:

1000 * (1 + .05/12) ***a*^b** (12 * 30)

Note: ***a*^b** is in the first row, second column of the main menu above. Now you can round your final answer to the nearest cent.

Attributions

This chapter contains material taken from *Math in Society* (on OpenTextBookStore) by David Lippman, and is used under a CC Attribution-Share Alike 3.0 United States (CC BY-SA 3.0 US) license.

This chapter contains material taken from of *Math for the Liberal Arts* (on Lumen Learning) by Lumen Learning, and is used under a *CC BY: Attribution* license.

Media Attributions

- Desmos Exponent Entry

- Compound Interest

ANNUITIES AND LOANS

Learning Outcomes

- Calculate the balance on an annuity after a specific amount of time
- Discern between compound interest, annuity, and payout annuity given a finance scenario
- Use the loan formula to calculate loan payments, loan balance, or interest accrued on a loan
- Determine which equation to use for a given scenario
- Solve a financial application for time

For most of us, we aren't able to put a large sum of money in the bank today. Instead, we save for the future by depositing a smaller amount of money from each paycheck into the bank. In this section, we will explore the math behind specific kinds of accounts that gain interest over time, like retirement accounts. We will also explore how mortgages and car loans, called installment loans, are calculated.

Savings Annuities

For most of us, we aren't able to put a large sum of money in the bank today. Instead, we save for the future by depositing a smaller amount of money from each paycheck into the bank. This idea is called a **savings annuity**. Most retirement plans like 401k plans or IRA plans are examples of savings annuities.

An annuity can be described recursively in a fairly simple way. Recall that basic compound interest follows from the relationship

$$P_m = \left(1 + \frac{r}{k}\right) P_{m-1}$$

For a savings annuity, we simply need to add a deposit, d, to the account with each compounding period:

$$P_m = \left(1 + \frac{r}{k}\right) P_{m-1} + d$$

Taking this equation from recursive form to explicit form is a bit trickier than with compound interest. It will be easiest to see by working with an example rather than working in general.

Example

Suppose we will deposit \$100 each month into an account paying 6% interest. We assume that the account is compounded with the same frequency as we make deposits unless stated otherwise. Write an explicit formula that represents this scenario.

Solution:

In this example:

- *r* = 0.06 (6%)
- *k* = 12 (12 compounds/deposits per year)
- *d* = \$100 (our deposit per month)

Writing out the recursive equation gives

$$P_m = \left(1 + \frac{0.06}{12}\right) P_{m-1} + 100 = (1.005)\, P_{m-1} + 100$$

Assuming we start with an empty account, we can begin using this relationship:

$$P_0 = 0$$

$$P_1 = (1.005)P_0 + 100 = 100$$

$$P_2 = (1.005)P_1 + 100 = (1.005)(100) + 100 = 100(1.005) + 100$$

$$P_3 = (1.005)P_2 + 100 = (1.005)(100(1.005) + 100) + 100 = 100(1.005)^2 + 100(1.005) + 100$$

Continuing this pattern, after *m* deposits, we'd have saved:

$$P_m = 100(1.005)^{m-1} + 100(1.005)^{m-2} + L + 100(1.005) + 100$$

In other words, after *m* months, the first deposit will have earned compound interest for *m*-1 months. The second deposit will have earned interest for *m*-2 months. The last month's deposit (L) would have earned only one month's worth of interest. The most recent deposit will have earned no interest yet.

This equation leaves a lot to be desired, though – it doesn't make calculating the ending balance any easier! To simplify things, multiply both sides of the equation by 1.005:

$$1.005P_m = 1.005\left(100(1.005)^{m-1} + 100(1.005)^{m-2} + \cdots + 100(1.005) + 100\right)$$

Distributing on the right side of the equation gives

$$1.005P_m = 100(1.005)^m + 100(1.005)^{m-1} + \cdots + 100(1.005)^2 + 100(1.005)$$

Now we'll line this up with like terms from our original equation, and subtract each side

(1)

$$\begin{array}{lclllll} 1.005P_m & = & 100(1.005)^m + & 100(1.005)^{m-1} & + \cdots + & 100(1.005) & \\ P_m & = & & 100(1.005)^{m-1} & + \cdots + & 100(1.005) & +100 \end{array}$$

Almost all the terms cancel on the right hand side when we subtract, leaving

$$1.005P_m - P_m = 100(1.005)^m - 100$$

Factor P_m out of the terms on the left side.

$$P_m(1.005 - 1) = 100(1.005)^m - 100$$
$$(0.005)P_m = 100(1.005)^m - 100$$

Solve for P_m

$$0.005P_m = 100\left((1.005)^m - 1\right)$$

(2) $$P_m = \frac{100\left((1.005)^m - 1\right)}{0.005}$$

Replacing *m* months with 12*N*, where *N* is measured in years, gives

$$P_N = \frac{100\left((1.005)^{12N} - 1\right)}{0.005}$$

Recall 0.005 was *r/k* and 100 was the deposit *d.* 12 was *k*, the number of deposit each year.

Generalizing this result, we get the savings annuity formula.

Annuity Formula

$$P_N = \frac{d\left(\left(1+\frac{r}{k}\right)^{Nk}-1\right)}{\left(\frac{r}{k}\right)}$$

- P_N is the balance in the account after N years.
- d is the regular deposit (the amount you deposit each year, each month, etc.)
- r is the annual interest rate in decimal form.
- k is the number of compounding periods in one year.

If the compounding frequency is not explicitly stated, assume there are the same number of compounds in a year as there are deposits made in a year.

For example, if the compounding frequency isn't stated:

- If you make your deposits every month, use monthly compounding, $k = 12$.
- If you make your deposits every year, use yearly compounding, $k = 1$.
- If you make your deposits every quarter, use quarterly

compounding, $k = 4$.

- Etc.

When do you use this?

Annuities assume that you put money in the account **on a regular schedule** (every month, year, quarter, etc.) and let it sit there earning interest.

Compound interest assumes that you put money in the account **once** and let it sit there earning interest.

- Compound interest: One deposit
- Annuity: Many deposits.

Examples

A traditional individual retirement account (IRA) is a special type of retirement account in which the money you invest is exempt from income taxes until you withdraw it. If you deposit $100 each month into an IRA earning 6% interest, how much will you have in the account after 20 years?

Solution:

In this example,

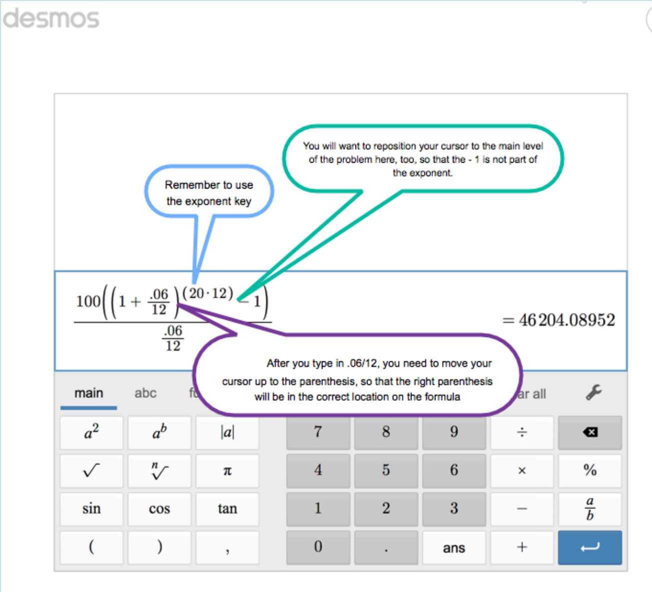
d = $100	the monthly deposit
r = 0.06	6% annual rate
k = 12	since we're doing monthly deposits, we'll compound monthly
N = 20	we want the amount after 20 years

Putting this into the equation:

$$P_N = \frac{d\left(\left(1+\frac{r}{k}\right)^{Nk}-1\right)}{\left(\frac{r}{k}\right)}$$

$$P_{20} = \frac{100\left(\left(1+\frac{0.06}{12}\right)^{240}-1\right)}{\left(\frac{0.05}{12}\right)}$$

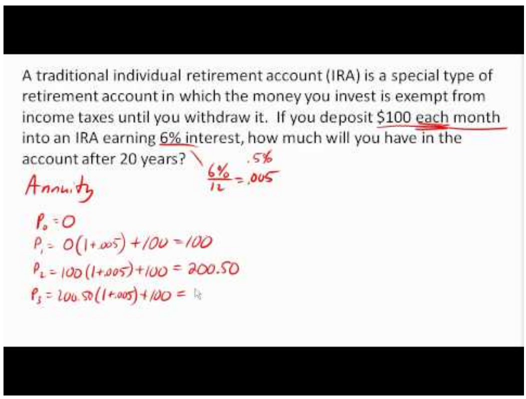
(Notice we multiplied N times k before putting it into the exponent. It is a simple computation and will make it easier to enter into Desmos:



The account will grow to $46,204.09 after 20 years.

Notice that you deposited into the account a total of $24,000 ($100 a month for 240 months). The difference between what you end up with and how much you put in is the interest earned. In this case it is $46,204.09 – $24,000 = $22,204.09.

This example is explained in detail here. Notice that each part was worked out separately and rounded. The answer above where we used Desmos is more accurate as the rounding was left until the end. You can work the problem either way, but be sure if you do follow the video below that you round out far enough for an accurate answer.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=135

Try It

A conservative investment account pays 3% interest.

If you deposit $5 a day into this account, how much will you have after 10 years? How much is from interest?

Solution:

d = $5 the daily deposit

r = 0.03 3% annual rate

k = 365 since we're doing daily deposits, we'll compound daily

N = 10 we want the amount after 10 years

$$P_{10} = \frac{5\left(\left(1+\frac{0.03}{365}\right)^{365*10}-1\right)}{\frac{0.03}{365}} = 21,282.07$$

Try It

Click here to try this problem.

Financial planners typically recommend that you have a certain amount of savings upon retirement. If you know the future value of the account, you can solve for the monthly contribution amount that will give you the desired result. In the next example, we will show you how this works.

Example

You want to have $200,000 in your account when you retire in 30 years. Your retirement account earns 8% interest. How much do you need to deposit each month to meet your retirement goal?
[reveal-answer q="897790"]Show Solution[/reveal-answer]
[hidden-answer a="897790"]

In this example, we're looking for *d*.

$r = 0.08$	8% annual rate
$k = 12$	since we're depositing monthly
$N = 30$	30 years
$P30 = \$200{,}000$	The amount we want to have in 30 years

In this case, we're going to have to set up the equation, and solve for *d*.

$$200,000 = \frac{d\left(\left(1+\frac{0.08}{12}\right)^{30(12)}-1\right)}{\left(\frac{0.08}{12}\right)}$$

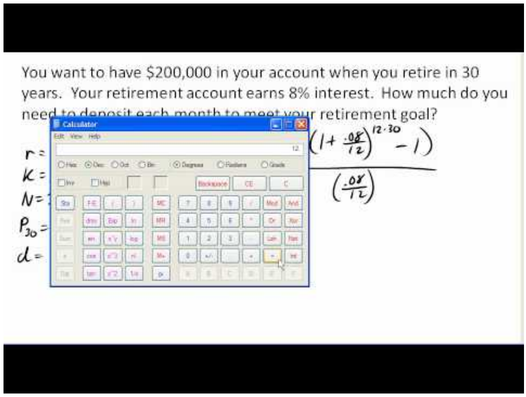
$$200,000 = \frac{d\left((1.00667)^{360}-1\right)}{(0.00667)}$$

$$200,000 = d(1491.57)$$

$$d = \frac{200,000}{1491.57} = \$134.09 \tag{3}$$

So you would need to deposit $134.09 each month to have $200,000 in 30 years if your account earns 8% interest.

View the solving of this problem in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=135

Try It

[Click here to try this problem.](#)

Payout Annuities

Removing Money from Annuities

In the last section you learned about annuities. In an annuity, you start with nothing, put money into an account on a regular basis, and end up with money in your account.

In this section, we will learn about a variation called a **Payout Annuity**. With a payout annuity, you start with money in the account, and pull money out of the account on a regular basis. Any remaining money in the account earns interest. After a fixed amount of time, the account will end up empty.

Payout annuities are typically used after retirement. Perhaps you have saved $500,000 for retirement, and want to take money out of the account each month to live on. You want the money to last you 20 years. This is a payout annuity. The formula is derived in a similar way as we did for savings annuities. The details are omitted here.

Payout Annuity Formula

$$P_0 = \frac{d\left(1-\left(1+\frac{r}{k}\right)^{-Nk}\right)}{\left(\frac{r}{k}\right)}$$

- P_0 is the balance in the account at the beginning (starting amount, or principal).
- d is the regular withdrawal (the amount you take out each year, each month, etc.)
- r is the annual interest rate (in decimal form. Example: 5% = 0.05)
- k is the number of compounding periods in one year.
- N is the number of years we plan to take withdrawals

Like with annuities, the compounding frequency is not always explicitly given, but is determined by how often you take the withdrawals.

When do you use this?

Payout annuities assume that you take money from the account on a regular schedule (every month, year, quarter, etc.) and let the rest sit there earning interest.

- Compound interest: One deposit
- Annuity: Many deposits.
- Payout Annuity: Many withdrawals

Example

After retiring, you want to be able to take $1000 every month for a total of 20 years from your retirement account. The account earns 6% interest. How much will you need in your account when you retire?
[reveal-answer q=”261541"]Show Solution[/reveal-answer]
[hidden-answer a=”261541"]

In this example,

$d =$ \$1000	the monthly withdrawal
$r =$ 0.06	6% annual rate
$k = 12$	since we're doing monthly withdrawals, we'll compound monthly
$N = 20$	since were taking withdrawals for 20 years

We're looking for P_0: how much money needs to be in the account at the beginning.

Putting this into the equation:

$$P_0 = \frac{1000\left(1-\left(1+\frac{0.06}{12}\right)^{-20(12)}\right)}{\left(\frac{0.06}{12}\right)}$$

$$P_0 = \frac{1000 \times \left(1-(1.005)^{-240}\right)}{(0.005)}$$

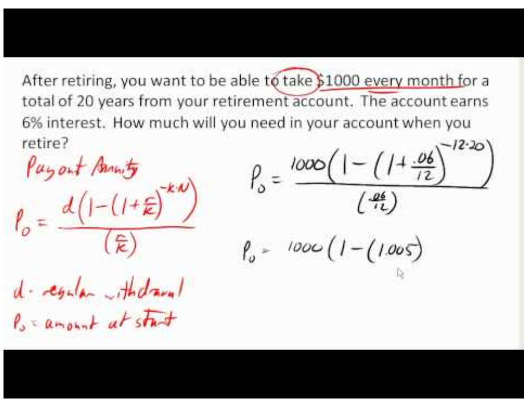
$$P_0 = \frac{1000 \times (1-0.302)}{(0.005)} = \$139,600 \tag{4}$$

You will need to have \$139,600 in your account when you retire.

The problem above was worked in sections, but remember you can entire the entire problem all at once in your Desmos calculator and avoid rounding.

Notice that you withdrew a total of \$240,000 (\$1000 a month for 240 months). The difference between what you pulled out and what you started with is the interest earned. In this case it is \$240,000 – \$139,600 = \$100,400 in interest.

View more about this problem in this video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=135

Try It

Click here to try this problem.

Evaluating negative exponents on your calculator

With these problems, you need to raise numbers to negative powers. Most calculators have a separate button for negating a number that is different than the subtraction button. Some calculators label this (-) , some with +/- . The button is often near the = key or the decimal point.

If your calculator displays operations on it (typically a calculator with multiline display), to calculate 1.005^{-240} you'd type something like: 1.005 ^ (-) 240

If your calculator only shows one value at a time, then usually you hit the (-) key after a number to negate it, so you'd hit: 1.005 yx 240 (-) =

Give it a try – you should get 1.005^{-240} = 0.302096

Example

You know you will have $500,000 in your account when you retire. You want to be able to take monthly withdrawals from the account for a total of 30 years. Your retirement account earns 8% interest. How much will you be able to withdraw each month?
[reveal-answer q="494776"]Show Solution[/reveal-answer]
[hidden-answer a="494776"]

In this example, we're looking for *d*.

$r = 0.08$	8% annual rate
$k = 12$	since we're withdrawing monthly
$N = 30$	30 years
$P_0 = \$500{,}000$	we are beginning with \$500,000

In this case, we're going to have to set up the equation, and solve for *d*.

$$500,000 = \frac{d\left(1-\left(1+\frac{0.08}{12}\right)^{-30(12)}\right)}{\left(\frac{0.08}{12}\right)}$$

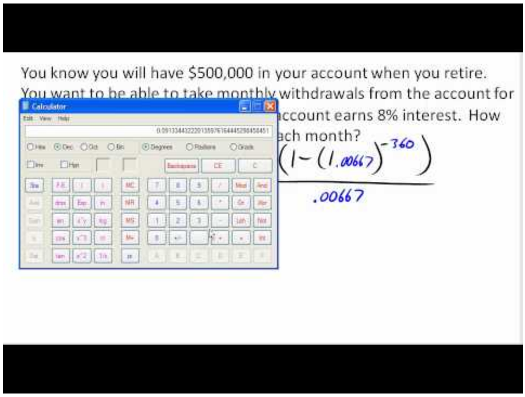
$$500,000 = \frac{d\left(1-(1.00667)^{-360}\right)}{(0.00667)}$$

$$500,000 = d(136.232)$$

$$d = \frac{500,000}{136.232} = \$3670.21 \tag{5}$$

You would be able to withdraw \$3,670.21 each month for 30 years.

A detailed walkthrough of this example can be viewed here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=135

Try It

[Click here to try this problem.]()

Try It

A donor gives $100,000 to a university, and specifies that it is to be used to give annual scholarships for the next 20 years. If the university can earn 4% interest, how much can they give in scholarships each year?

d = unknown

r = 0.04 4% annual rate

k = 1 since we're doing annual scholarships

N = 20 20 years

P0 = 100,000 we're starting with $100,000

$$100,000 = \frac{d\left(1-\left(1+\frac{0.04}{1}\right)^{-20*1}\right)}{\frac{0.04}{1}}$$

Solving for *d* gives $7,358.18 each year that they can give in scholarships.

It is worth noting that usually donors instead specify that only interest is to be used for scholarship, which makes the original donation last indefinitely. If this donor had specified that, $100,000(0.04) = $4,000 a year would have been available.

Loans

Conventional Loans

In the last section, you learned about payout annuities. In this section, you will learn about conventional loans (also called amortized loans or installment loans). Examples include auto loans and home mortgages. These techniques do not apply to payday loans, add-on loans, or other loan types where the interest is calculated up front.

One great thing about loans is that they use exactly the same formula as a payout annuity. To see why, imagine that you had $10,000 invested at a bank, and started taking out payments while earning interest as part of a payout annuity, and after 5 years your balance was zero. Flip that around, and imagine that you are acting as the bank, and a car lender is acting as you. The car lender invests $10,000 in you. Since you're acting as the bank, you pay interest. The car lender takes payments until the balance is zero.

Loans Formula

$$P_0 = \frac{d\left(1-\left(1+\frac{r}{k}\right)^{-Nk}\right)}{\left(\frac{r}{k}\right)}$$

- P_0 is the balance in the account at the beginning (the principal, or amount of the loan).
- d is your loan payment (your monthly payment, annual payment, etc)
- r is the annual interest rate in decimal form.
- k is the number of compounding periods in one year.
- N is the length of the loan, in years.

Like before, the compounding frequency is not always explicitly given, but is determined by how often you make payments.

When do you use this?

The loan formula assumes that you make loan payments on a regular schedule (every month, year, quarter, etc.) and are paying interest on the loan.

- Compound interest: One deposit
- Annuity: Many deposits
- Payout Annuity: Many withdrawals
- Loans: Many payments

Example

You can afford $200 per month as a car payment. If you can get an auto loan at 3% interest for 60 months (5 years), how expensive of a car can you afford? In other words, what amount loan can you pay off with $200 per month?

[reveal-answer q="129373"]Show Solution[/reveal-

answer]
[hidden-answer a=”129373"]

In this example,

d = \$200	the monthly loan payment
r = 0.03	3% annual rate
k = 12	since we're doing monthly payments, we'll compound monthly
N = 5	since we're making monthly payments for 5 years

We're looking for P_0, the starting amount of the loan.

$$P_0 = \frac{200\left(1-\left(1+\frac{0.03}{12}\right)^{-5(12)}\right)}{\left(\frac{0.03}{12}\right)}$$

$$P_0 = \frac{200\left(1-(1.0025)^{-60}\right)}{(0.0025)}$$

(6) $$P_0 = \frac{200\,(1-0.861)}{(0.0025)} = \$11,120$$

You can afford a \$11,120 loan.

You will pay a total of \$12,000 (\$200 per month for

60 months) to the loan company. The difference between the amount you pay and the amount of the loan is the interest paid. In this case, you're paying $12,000-$11,120 = $880 interest total.

Details of this example are examined in this video.

Click here to view this video.

Try It

Click here to try this problem.

Example

You want to take out a $140,000 mortgage (home loan). The interest rate on the loan is 6%, and the loan is for 30 years. How much will your monthly payments be?

Solution:

In this example, we're looking for *d*.

$r = 0.06$	6% annual rate
$k = 12$	since we're paying monthly
$N = 30$	30 years
$P_0 = \$140{,}000$	the starting loan amount

In this case, we're going to have to set up the equation, and solve for *d*.

$$140,000 = \frac{d\left(1 - \left(1 + \frac{0.06}{12}\right)^{-30(12)}\right)}{\left(\frac{0.06}{12}\right)}$$

$$140,000 = \frac{d\left(1 - (1.005)^{-360}\right)}{(0.005)}$$

$$140,000 = d(166.792)$$

$$d = \frac{140,000}{166.792} = \$839.37 \tag{7}$$

You will make payments of $839.37 per month for 30 years.

You're paying a total of $302,173.20 to the loan company: $839.37 per month for 360 months. You are paying a total of $302,173.20 – $140,000 = $162,173.20 in interest over the life of the loan.

View more about this example here.

Try It

Click here to try this problem.

Try It

Janine bought $3,000 of new furniture on credit. Because her credit score isn't very good, the store is charging her a fairly high interest rate on the loan: 16%. If she agreed to pay off the furniture over 2 years, how much will she have to pay each month?

Solution:

d = unknown

r = 0.16 16% annual rate

k = 12 since we're making monthly payments

N = 2 2 years to repay

P0 = 3,000 we're starting with a $3,000 loan

$$3000=\frac{d\left(1-\left(1+\frac{0.06}{12}\right)^{-2*12}\right)}{\frac{0.16}{12}}$$

$$3000=20.42d$$

Solve for d to get monthly payments of $146.89

Two years to repay means $146.89(24) = $3525.36 in total payments. This means Janine will pay $3525.36 – $3000 = $525.36 in interest.

Calculating the Balance

With loans, it is often desirable to determine what the remaining loan balance will be after some number of years. For example, if you purchase a home and plan to sell it in five years, you might want to know how much of the loan balance you will have paid off and how much you have to pay from the sale.

To determine the remaining loan balance after some number of years, we first need to know the loan payments, if we don't already know them. Remember that only a portion of your loan payments go towards the loan balance; a portion is going to go towards interest. For example, if your payments were \$1,000 a month, after a year you will *not* have paid off \$12,000 of the loan balance.

To determine the remaining loan balance, we can think "how much loan will these loan payments be able to pay off in the remaining time on the loan?"

Example

If a mortgage at a 6% interest rate has payments of \$1,000 a month, how much will the loan balance be 10 years from the end the loan?

Solution:

To determine this, we are looking for the amount of the loan that can be paid off by \$1,000 a month payments in 10 years. In other words, we're looking for P_0 when

$d =$ \$1,000	the monthly loan payment
$r = 0.06$	6% annual rate
$k = 12$	since we're doing monthly payments, we'll compound monthly
$N = 10$	since we're making monthly payments for 10 more years

$$P_0 = \frac{1000\left(1-\left(1+\frac{0.06}{12}\right)^{-10(12)}\right)}{\left(\frac{0.06}{12}\right)}$$

$$P_0 = \frac{1000\left(1-(1.005)^{-120}\right)}{(0.005)}$$

$$P_0 = \frac{1000\left(1-0.5496\right)}{(0.005)} = \$90,073.45 \tag{8}$$

The loan balance with 10 years remaining on the loan will be \$90,073.45.

This example is explained in this video:

Oftentimes answering remaining balance questions requires two steps:

1. Calculating the monthly payments on the loan
2. Calculating the remaining loan balance based on the *remaining time* on the loan

Example

A couple purchases a home with a \$180,000 mortgage at 4% for 30 years with monthly payments. What will the remaining balance on their mortgage be after 5 years?

Solution:

First we will calculate their monthly payments.

We're looking for *d*.

$r = 0.04$	4% annual rate
$k = 12$	since they're paying monthly
$N = 30$	30 years
$P_0 = \$180{,}000$	the starting loan amount

We set up the equation and solve for *d*.

$$180,000 = \frac{d\left(1-\left(1+\frac{0.04}{12}\right)^{-30(12)}\right)}{\left(\frac{0.04}{12}\right)}$$

$$180,000 = \frac{d\left(1-(1.00333)^{-360}\right)}{(0.00333)}$$

$$180,000 = d(209.562)$$

$$d = \frac{180,000}{209.562} = \$858.93 \tag{9}$$

Now that we know the monthly payments, we can determine the remaining balance. We want the remaining balance after 5 years, when 25 years will be remaining on the loan, so we calculate the loan balance that will be paid off with the monthly payments over those 25 years.

d = \$858.93	the monthly loan payment we calculated above
$r = 0.04$	4% annual rate
$k = 12$	since they're doing monthly payments
$N = 25$	since they'd be making monthly payments for 25 more years

(10)

$$P_0 = \frac{858.93\left(1-\left(1+\frac{0.04}{12}\right)^{-25(12)}\right)}{\left(\frac{0.04}{12}\right)}$$

$$P_0 = \frac{858.93\left(1-(1.00333)^{-300}\right)}{(0.00333)}$$

$$P_0 = \frac{858.93\,(1-0.369)}{(0.00333)} = \$155,793.91$$

The loan balance after 5 years, with 25 years remaining on the loan, will be $155,793.91.

Over that 5 years, the couple has paid off $180,000 – $155,793.91 = $24,206.09 of the loan balance. They have paid a total of $858.93 a month for 5 years (60 months), for a total of $51,535.80, so $51,535.80 – $24,206.09 = $27,329.71 of what they have paid so far has been interest.

More explanation of this example is available here:

FYI

Home loans are typically paid off through an amortization process, **amortization** refers to paying off a debt (often from a loan or mortgage) over time through regular payments. An **amortization schedule** is a **table** detailing each periodic payment on an **amortizing** loan as generated by an **amortization calculator**.

If you want to know more, click on the link below to view the website "How is an Amortization Schedule Calculated?" by MyAmortizationChart.com. This website provides a brief overlook of Amortization Schedules.

- How is an Amortization Schedule Calculated?

Click here to try this problem.

Attributions

Media Attributions

- Desmos Annuity.1

WHICH EQUATION TO USE?

Which Formula to Use?

Now that we have surveyed the basic kinds of finance calculations that are used, it may not always be obvious which one to use when you are given a problem to solve. Here are some hints on deciding which equation to use, based on the wording of the problem.

Loans

The easiest types of problems to identify are loans. Loan problems almost always include words like **loan**, **amortize** (the fancy word for loans), **finance** (i.e. a car), or **mortgage** (a home loan). Look for words like monthly or annual payment.

The loan formula assumes that you make loan payments on a regular schedule (every month, year, quarter, etc.) and are paying interest on the loan.

Loans Formula

$$P_0 = \frac{d\left(1-\left(1+\frac{r}{k}\right)^{-Nk}\right)}{\left(\frac{r}{k}\right)}$$

- P_0 is the balance in the account at the beginning (the principal, or amount of the loan).
- d is your loan payment (your monthly payment, annual payment, etc)
- r is the annual interest rate in decimal form.
- k is the number of compounding periods in one year.
- N is the length of the loan, in years.

Interest-Bearing Accounts

Accounts that gain interest fall into two main categories. The first is on where you put money in an account once and let it sit, the other is where you make regular payments or withdrawals from the account as in a retirement account.

Interest

- If you're letting the money sit in the account with nothing but interest changing the balance, then you're looking at a **compound interest** problem. Look for words like compounded, or APY. Compound interest assumes that you put money in the account **once** and let it sit there earning interest.

COMPOUND INTEREST

$$P_N = P_0 \left(1 + \frac{r}{k}\right)^{Nk}$$

- P_N is the balance in the account after N years.

- P_0 is the starting balance of the account (also called initial deposit, or principal)
- r is the annual interest rate in decimal form
- k is the number of compounding periods in one year
    - If the compounding is done annually (once a year), $k = 1$.
    - If the compounding is done quarterly, $k = 4$.
    - If the compounding is done monthly, $k = 12$.
    - If the compounding is done daily, $k = 365$.

- The exception would be bonds and other investments where the interest is not reinvested; in those cases you're looking at **simple interest**.

SIMPLE INTEREST OVER TIME

(1)

$$I = P_0rt$$

$$A = P_0 + I = P_0 + P_0rt = P_0(1 + rt)$$

- *I* is the interest
- *A* is the end amount: principal plus interest
- (2) P_0

  is the principal (starting amount)
- *r* is the interest rate in decimal form
- *t* is time

The units of measurement (years, months, etc.) for the time should match the time period for the interest rate.

Annuities

- If you're putting money ***into*** the account on a regular basis (monthly/annually/quarterly) then you're looking at a **basic annuity** problem. Basic annuities are when you are saving money. Usually in an annuity problem, your account starts empty, and has money in the future. Annuities assume that you put money in the account **on a regular schedule** (every month, year, quarter, etc.) and let it sit there earning interest.

ANNUITY FORMULA

$$P_N = \frac{d\left(\left(1+\frac{r}{k}\right)^{Nk}-1\right)}{\left(\frac{r}{k}\right)}$$

- P_N is the balance in the account after N years.
- d is the regular deposit (the amount you deposit each year, each month, etc.)
- r is the annual interest rate in decimal form.
- k is the number of compounding periods in one year.

If the compounding frequency is not explicitly stated, assume there are the same number of compounds in a year as there are deposits made in a year.

- If you're pulling money ***out*** of the account on a regular basis, then you're looking at a **payout annuity** problem. Payout annuities are used for things like retirement income, where you start with money in your account, pull money out on a regular basis, and your account ends up empty in the future. Payout annuities assume that you take money from the account on a regular schedule (every month, year, quarter, etc.) and let the rest sit there earning interest.

PAYOUT ANNUITY FORMULA

$$P_0 = \frac{d\left(1-\left(1+\frac{r}{k}\right)^{-Nk}\right)}{\left(\frac{r}{k}\right)}$$

- P_0 is the balance in the account at the beginning (starting amount, or principal).
- d is the regular withdrawal (the amount you take out each year, each month, etc.)
- r is the annual interest rate (in decimal form. Example: 5% = 0.05)
- k is the number of compounding periods in one year.
- N is the number of years we plan to take

withdrawals

Remember, the most important part of answering any kind of question, money or otherwise, is first to correctly identify what the question is really asking, and then determine what approach will best allow you to solve the problem.

Try It

For each of the following scenarios, determine if it is a compound interest problem, a savings annuity problem, a payout annuity problem, or a loans problem. Then solve each problem.

1. Marcy received an inheritance of $20,000, and invested it at 6% interest. She is going to use it for college, withdrawing money for tuition and expenses each quarter. How much can she take out each quarter if she has 3 years of school left?
2. Paul wants to buy a new car. Rather than take

out a loan, he decides to save $200 a month in an account earning 3% interest compounded monthly. How much will he have saved up after 3 years?

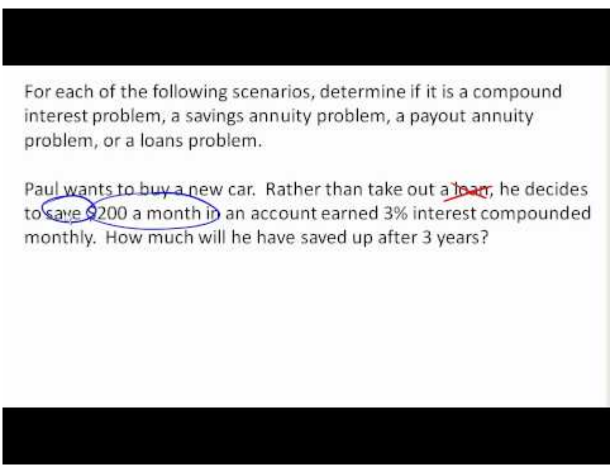
3. Keisha is managing investments for a non-profit company. They want to invest some money in an account earning 5% interest compounded annually with the goal to have $30,000 in the account in 6 years. How much should Keisha deposit into the account?
4. Miao is going to finance new office equipment at a 2% rate over a 4 year term. If she can afford monthly payments of $100, how much new equipment can she buy?
5. How much would you need to save every month in an account earning 4% interest to have $5,000 saved up in two years?

Solutions:

1. This is a payout annuity problem. She can pull out $1833.60 a quarter.
2. This is a savings annuity problem. He will have saved up $7,524.11
3. This is compound interest problem. She would need to deposit $22,386.46.

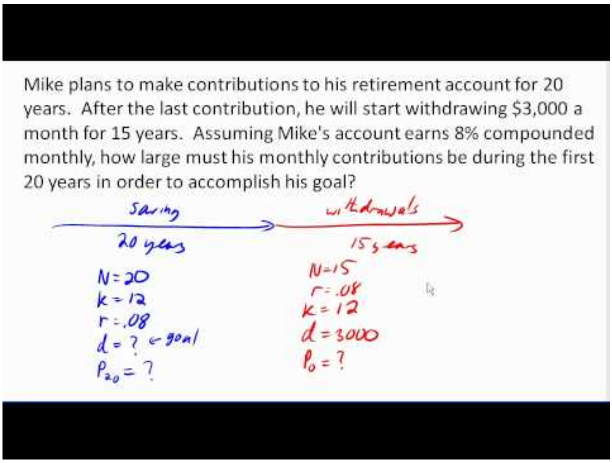
4. This is a loans problem. She can buy \$4,609.33 of new equipment
5. This is a savings annuity problem. You would need to save \$200.46 each month

In the following video, we present more examples of how to use the language in the question to determine which type of equation to use to solve a finance problem.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=379

In the next video example, we show how to solve a finance problem that has two stages, the first stage is a savings problem, and the second stage is a withdrawal problem.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=379

Try It

[Click here to try this problem.]()

Try It

Click here to try this problem.

Try It

Click here to try this problem.

Attributions

PUTTING IT TOGETHER: FINANCE

Financing a Refrigerator: Three Scenarios

In the beginning of this module, we presented three options for buying a new refrigerator. In one scenario, you could rent to own, with the following terms $17.99 per week for 2 years, which is 104 weeks. The total cost is:

$$104 \times 17.99 = 1870.96$$

Scenario two involved a loan of $1299 from your brother at 20% interest for one full year. To calculate the total amount, use the **simple interest formula**,

$$I = P_0 rt$$

In this situation, the principle amount is $P_0 = 1299$, rate is $r = 20\% = 0.20$, and the time is $t = 1$ year. Therefore, the interest due to your brother would be:

$$I = 1299 \times 0.20 \times 1 = 259.80$$

Adding the interest back to the principle, the total cost of the refrigerator would amount to $1558.80. That's quite a bit less than the $1870.96 that the rent-to-own store would ultimately have received from you. But your brother wants the money in one year, so let's figure out what the weekly payment would be. Simply divide the total by 52 weeks.

$$1558.80 \div 52 = 29.98$$

This is a higher weekly payment than the rent-to-own store is offering, but if you can afford it, then you'll save money in the long run.

Finally, let's explore the third option. This time we use the **loans formula**,

$$P_0 = \frac{d\left(1-\left(1+\frac{r}{k}\right)^{-Nk}\right)}{\left(\frac{r}{k}\right)}$$

The principle is the same, $P_0 = 1299$, but the rate is now $r = 15\% = 0.15$. Because the compounding is monthly, we have $k = 12$. Finally, $N = 3$ represents the total number of years for the loan. We must solve for d.

$$1299 = \frac{d\left(1-\left(1+\frac{0.15}{12}\right)^{-3(12)}\right)}{\left(\frac{0.15}{12}\right)}$$
$$1299 = \frac{d\left(1-(1.0125)^{-36}\right)}{0.0125}$$
$$1299 = \frac{d(0.36059)}{0.0125}$$
$$d = 1299 \times 0.0125 \div 0.36059 = 45.03$$

This calculation gives the monthly payment (since the compounding is monthly), $d = 45.03$ If we want to see how this compares against our previous scenarios, we can find an equivalent weekly payment. The best way to do this is to multiply d by 12 and then divide by 52. This gives a weekly payment of about $45.03 \times 12 \div 52 = 10.39$, by far the lowest weekly payment, but what is the total cost?

Finally, to calculate the total cost, multiply the monthly payment by the number of months in 3 years, that is, 36 months.

$$45.03 \times 36 = 1621.09$$

Option three, the line of store credit. This option seemed pretty good at first. However, because of the long loan period and compounding interest, the total cost is actually more than the $1558.80 from scenario two.

Let's compare the details of each scenario shown in the table below. Note, the **total interest** is found by subtracting the list price of the refrigerator ($1299) from the **total paid** amount.

	Rent to Own	Brother's Offer	Store Loan
Payments	\$17.99 per week	\$29.98 per week	\$45.03 per month (\$10.39 per week)
Length of Term	2 years	1 year	3 years
Total Paid	\$1870.96	\$1558.80	\$1621.09
Total Interest	\$571.96	\$259.80	\$322.09

Which scenario would you choose?

Attributions

EXERCISES

Skills

1. A friend lends you $200 for a week, which you agree to repay with 5% one-time interest. How much will you have to repay?
2. Suppose you obtain a $3,000 T-note with a 3% annual rate, paid quarterly, with maturity in 5 years. How much interest will you earn?
3. A T-bill is a type of bond that is sold at a discount over the face value. For example, suppose you buy a 13-week T-bill with a face value of $10,000 for $9,800. This means that in 13 weeks, the government will give you the face value, earning you $200. What annual interest rate have you earned?
4. Suppose you are looking to buy a $5000 face value 26-week T-bill. If you want to earn at least 1% annual interest, what is the most you should pay for the T-bill?
5. You deposit $300 in an account earning 5% interest compounded annually. How much will you have in the account in 10 years?
6. How much will $1000 deposited in an account earning 7% interest compounded annually be worth in 20 years?

7. You deposit \$2000 in an account earning 3% interest compounded monthly.
   a. How much will you have in the account in 20 years?
   b. How much interest will you earn?
8. You deposit \$10,000 in an account earning 4% interest compounded monthly.
   a. How much will you have in the account in 25 years?
   b. How much interest will you earn?
9. How much would you need to deposit in an account now in order to have \$6,000 in the account in 8 years? Assume the account earns 6% interest compounded monthly.
10. How much would you need to deposit in an account now in order to have \$20,000 in the account in 4 years? Assume the account earns 5% interest.
11. You deposit \$200 each month into an account earning 3% interest compounded monthly.
    a. How much will you have in the account in 30 years?
    b. How much total money will you put into the account?
    c. How much total interest will you earn?
12. You deposit \$1000 each year into an account earning 8% compounded annually.
    a. How much will you have in the account in 10

years?

b. How much total money will you put into the account?
c. How much total interest will you earn?

13. Jose has determined he needs to have $800,000 for retirement in 30 years. His account earns 6% interest.
    a. How much would you need to deposit in the account each month?
    b. How much total money will you put into the account?
    c. How much total interest will you earn?
14. You wish to have $3000 in 2 years to buy a fancy new stereo system. How much should you deposit each quarter into an account paying 8% compounded quarterly?
15. You want to be able to withdraw $30,000 each year for 25 years. Your account earns 8% interest.
    a. How much do you need in your account at the beginning
    b. How much total money will you pull out of the account?
    c. How much of that money is interest?
16. How much money will I need to have at retirement so I can withdraw $60,000 a year for 20 years from an account earning 8% compounded annually?
    a. How much do you need in your account at the beginning

   b. How much total money will you pull out of the account?
   c. How much of that money is interest?

17. You have $500,000 saved for retirement. Your account earns 6% interest. How much will you be able to pull out each month, if you want to be able to take withdrawals for 20 years?
18. Loren already knows that he will have $500,000 when he retires. If he sets up a payout annuity for 30 years in an account paying 10% interest, how much could the annuity provide each month?
19. You can afford a $700 per month mortgage payment. You've found a 30 year loan at 5% interest.
    a. How big of a loan can you afford?
    b. How much total money will you pay the loan company?
    c. How much of that money is interest?
20. Marie can afford a $250 per month car payment. She's found a 5 year loan at 7% interest.
    a. How expensive of a car can she afford?
    b. How much total money will she pay the loan company?
    c. How much of that money is interest?
21. You want to buy a $25,000 car. The company is offering a 2% interest rate for 48 months (4 years). What will your monthly payments be?
22. You decide finance a $12,000 car at 3% compounded

monthly for 4 years. What will your monthly payments be? How much interest will you pay over the life of the loan?

23. You want to buy a $200,000 home. You plan to pay 10% as a down payment, and take out a 30 year loan for the rest.
    a. How much is the loan amount going to be?
    b. What will your monthly payments be if the interest rate is 5%?
    c. What will your monthly payments be if the interest rate is 6%?
24. Lynn bought a $300,000 house, paying 10% down, and financing the rest at 6% interest for 30 years.
    a. Find her monthly payments.
    b. How much interest will she pay over the life of the loan?
25. Emile bought a car for $24,000 three years ago. The loan had a 5 year term at 3% interest rate. How much does he still owe on the car?
26. A friend bought a house 15 years ago, taking out a $120,000 mortgage at 6% for 30 years. How much does she still owe on the mortgage?
27. Pat deposits $6,000 into an account earning 4% compounded monthly. How long will it take the account to grow to $10,000?
28. Kay is saving $200 a month into an account earning 5% interest. How long will it take her to save $20,000?

29. James has $3,000 in credit card debt, which charges 14% interest. How long will it take to pay off the card if he makes the minimum payment of $60 a month?
30. Chris has saved $200,000 for retirement, and it is in an account earning 6% interest. If she withdraws $3,000 a month, how long will the money last?

Concepts

31. Suppose you invest $50 a month for 5 years into an account earning 8% compounded monthly. After 5 years, you leave the money, without making additional deposits, in the account for another 25 years. How much will you have in the end?
32. Suppose you put off making investments for the first 5 years, and instead made deposits of $50 a month for 25 years into an account earning 8% compounded monthly. How much will you have in the end?
33. Mike plans to make contributions to his retirement account for 15 years. After the last contribution, he will start withdrawing $10,000 a quarter for 10 years. Assuming Mike's account earns 8% compounded quarterly, how large must his quarterly contributions be during the first 15 years, in order to accomplish his goal?
34. Kendra wants to be able to make withdrawals of $60,000 a year for 30 years after retiring in 35 years. How much will she have to save each year up until retirement if her

account earns 7% interest?

35. You have $2,000 to invest, and want it to grow to $3,000 in two years. What interest rate would you need to find to make this possible?
36. You have $5,000 to invest, and want it to grow to $20,000 in ten years. What interest rate would you need to find to make this possible?
37. You plan to save $600 a month for the next 30 years for retirement. What interest rate would you need to have $1,000,000 at retirement?
38. You really want to buy a used car for $11,000, but can only afford $200 a month. What interest rate would you need to find to be able to afford the car, assuming the loan is for 60 months?

Exploration

39. Pay day loans are short term loans that you take out against future paychecks: The company advances you money against a future paycheck. Either visit a pay day loan company, or look one up online. Be forewarned that many companies do not make their fees obvious, so you might need to do some digging or look at several companies.
    a. Explain the general method by which the loan works.
    b. We will assume that we need to borrow $500 and

that we will pay back the loan in 14 days. Determine the total amount that you would need to pay back and the effective loan rate. The effective loan rate is the percentage of the original loan amount that you pay back. It is not the same as the APR (annual rate) that is probably published.

   c. If you cannot pay back the loan after 14 days, you will need to get an extension for another 14 days. Determine the fees for an extension, determine the total amount you will be paying for the now 28 day loan, and compute the effective loan rate.

40. Suppose that 10 years ago you bought a home for $110,000, paying 10% as a down payment, and financing the rest at 9% interest for 30 years.
    a. Let's consider your existing mortgage:
        i. How much money did you pay as your down payment?
        ii. How much money was your mortgage (loan) for?
        iii. What is your current monthly payment?
        iv. How much total interest will you pay over the life of the loan?
    b. This year, you check your loan balance. Only part of your payments have been going to pay down the loan; the rest has been going towards interest. You see that you still have $88,536 left to pay on your loan. Your house is now valued at $150,000.

i. How much of the loan have you paid off? (i.e., how much have you reduced the loan balance by? Keep in mind that interest is charged each month – it's not part of the loan balance.)
ii. How much money have you paid to the loan company so far?
iii. How much interest have you paid so far?
iv. How much equity do you have in your home (equity is value minus remaining debt)

c. Since interest rates have dropped, you consider refinancing your mortgage at a lower 6% rate.
   i. If you took out a new 30 year mortgage at 6% for your remaining loan balance, what would your new monthly payments be?
   ii. How much interest will you pay over the life of the new loan?

d. Notice that if you refinance, you are going to be making payments on your home for another 30 years. In addition to the 10 years you've already been paying, that's 40 years total.
   i. How much will you save each month because of the lower monthly payment?
   ii. How much total interest will you be paying (you need to consider the amount from 2c and 3b)
   iii. Does it make sense to refinance? (there isn't a correct answer to this question. Just give your

opinion and your reason)

Attributions

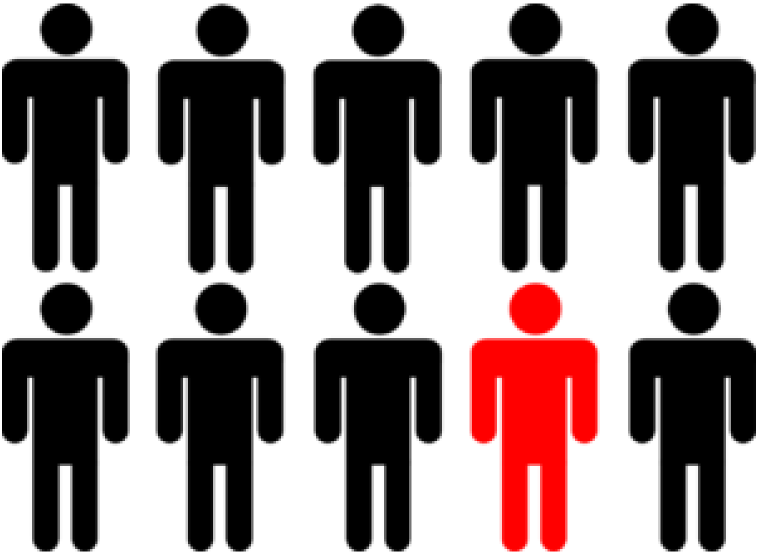
CHAPTER IV

CHAPTER 4: STATISTICS: COLLECTING DATA

WHY IT MATTERS: STATISTICS: COLLECTING DATA

Why understand how data is collected?

Like most people, you probably feel that it is important to "take control of your life." But what does this mean? Partly it means being able to properly evaluate the data and claims that bombard you every day. If you cannot distinguish good from faulty reasoning, then you are vulnerable to manipulation and to decisions that are not in your best interest. Statistics provides tools that you need in order to react intelligently to information you hear or read. In this sense, Statistics is one of the most important things that you can study.



To be more specific, here are some claims that we have heard on several occasions.

(We are *not* saying that each one of these claims is true!)

- 4 out of 5 dentists recommend Dentyne.
- Almost 85% of lung cancers in men and 45% in women are tobacco-related.
- Condoms are effective 94% of the time.
- Native Americans are significantly more likely to be hit crossing the streets than are people of other ethnicities.
- People tend to be more persuasive when they look others directly in the eye and speak loudly and quickly.
- Women make 75 cents to every dollar a man makes when they work the same job.

- A surprising new study shows that eating egg whites can increase one's life span.
- People predict that it is very unlikely there will ever be another baseball player with a batting average over 400.
- There is an 80% chance that in a room full of 30 people that at least two people will share the same birthday.
- 79.48% of all statistics are made up on the spot.

All of these claims are statistical in character. We suspect that some of them sound familiar; if not, we bet that you have heard other claims like them. Notice how diverse the examples are; they come from psychology, health, law, sports, business, etc. Indeed, data and data-interpretation show up in discourse from virtually every facet of contemporary life.

Statistics are often presented in an effort to add credibility to an argument or advice. You can see this by paying attention to television advertisements. Many of the numbers thrown about in this way do not represent careful statistical analysis. They can be misleading, and push you into decisions that you might find cause to regret. For these reasons, learning about statistics is a long step towards taking control of your life. (It is not, of course, the only step needed for this purpose.) These chapters will help you learn statistical essentials. It will make you into an intelligent consumer of statistical claims.

You can take the first step right away. To be an intelligent consumer of statistics, your first reflex must be to question the statistics that you encounter. The British Prime Minister Benjamin Disraeli famously said, “There are three kinds of lies—lies, damned lies, and statistics.” This quote reminds us why it is so important to understand statistics. So let us invite you to reform your statistical habits from now on. No longer will you blindly accept numbers or findings. Instead, you will begin to think about the numbers, their sources, and most importantly, the procedures used to generate them.

We have put the emphasis on defending ourselves against fraudulent claims wrapped up as statistics. Just as important as detecting the deceptive use of statistics is the appreciation of the proper use of statistics. You must also learn to recognize statistical evidence that supports a stated conclusion. When a research team is testing a new treatment for a disease, statistics allows them to conclude based on a relatively small trial that there is good evidence their drug is effective. Statistics allowed prosecutors in the 1950s and 60s to demonstrate racial bias existed in jury panels. Statistics are all around you, sometimes used well, sometimes not. We must learn how to distinguish the two cases.

Before we dive in, let's see a practical case of statistical analysis in action. You've likely seen several TED Talks videos at this point, either elsewhere in this course, in other courses you're taking, or out of personal interest. This amusing video analyzes the data that TED has gathered and produces some tips for how to put together the most (or least!) effective TED Talk possible.

A YouTube element has been excluded from this

version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=141

Attributions

DATA COLLECTION BASICS

Learning Outcomes

- Determine whether a value calculated from a group is a statistic or a parameter
- Identify the difference between a census and a sample
- Identify the population of a study
- Determine whether a measurement is categorical or qualitative

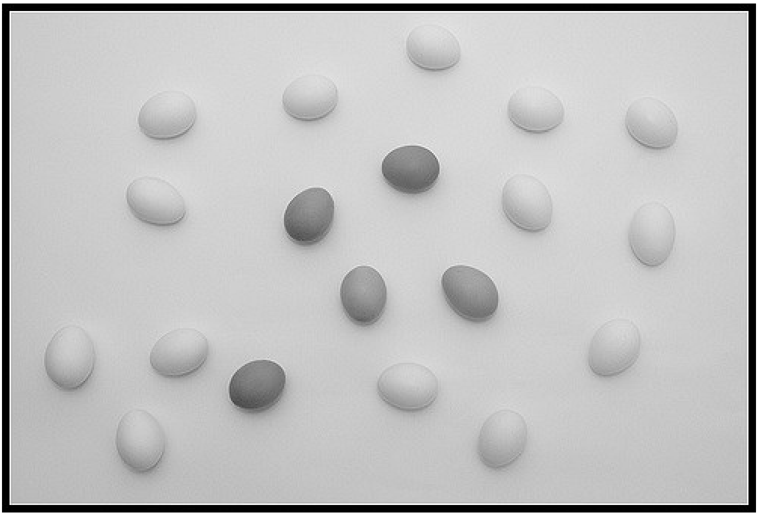
In this lesson we will introduce some important terminology related to collecting data. When you are finished you will be able to identify the difference between terms like census and sample. In the following lessons we will rely on your understanding of these terms, so study well!

Populations and Samples

Selecting A Focus

Before we begin gathering and analyzing data we need to characterize the **population** we are studying. If we want to study the amount of money spent on textbooks by a typical first-year college student, our population might be all first-year students at your college. Or it might be:

- All first-year community college students in the state of Washington.
- All first-year students at public colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the state of Washington.
- All first-year students at all colleges and universities in the entire United States.
- And so on.



Population

The **population** of a study is the group the collected data is intended to describe.

Sometimes the intended population is called the **target population**, since if we design our study badly, the collected data might not actually be representative of the intended population.

Why is it important to specify the population? We might get different answers to our question as we vary the population we are studying. First-year students at the University of Washington might take slightly more diverse courses than those at your college, and some of these courses may require less popular textbooks that cost more; or, on the other hand, the University Bookstore might have a larger pool of used textbooks, reducing the cost of these books to the students. Whichever the case (and it is likely that some combination of these and other factors are in play), the data we gather from your college will probably not be the same as that from the University of Washington. Particularly when conveying our results to others, we want to be clear about the population we are describing with our data.

example

A newspaper website contains a poll asking people their opinion on a recent news article.

What is the population?

Solution:

While the target (intended) population may have been all people, the real population of the survey is readers of the website.

If we were able to gather data on every member of our population, say the average (we will define "average" more carefully in a subsequent section) amount of money spent on textbooks by each first-year student at your college during the 2009-2010 academic year, the resulting number would be called a **parameter**.

Parameter

A **parameter** is a value (average, percentage, etc.) calculated using all the data from a population

We seldom see parameters, however, since surveying an entire population is usually very time-consuming and expensive, unless the population is very small or we already have the data collected.

Census

A survey of an entire population is called a **census**.

You are probably familiar with two common censuses: the official government Census that attempts to count the population of the U.S. every ten years, and voting, which asks the opinion of all eligible voters in a district. The first of these demonstrates one additional problem with a census: the difficulty in finding and getting participation from everyone in a large population, which can bias, or skew, the results.

There are occasionally times when a census is appropriate, usually when the population is fairly small. For example, if the manager of Starbucks wanted to know the average number of hours her employees worked last week, she should be able to pull up payroll records or ask each employee directly.

Since surveying an entire population is often impractical, we usually select a **sample** to study.

Sample

A **sample** is a smaller subset of the entire population, ideally one that is fairly representative of the whole population.

We will discuss sampling methods in greater detail in a later section. For now, let us assume that samples are chosen in an appropriate manner. If we survey a sample, say 100 first-year students at your college, and find the average amount of money spent by these students on textbooks, the resulting number is called a **statistic**.

Statistic

A **statistic** is a value (average, percentage, etc.) calculated using the data from a sample.

example

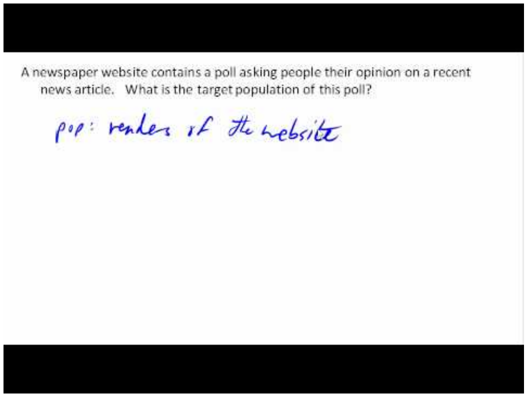
A researcher wanted to know how citizens of Tacoma felt about a voter initiative. To study this, she goes to the Tacoma Mall and randomly selects 500 shoppers and asks them their opinion. 60% indicate they are supportive of the initiative. What is the sample and population? Is the 60% value a parameter or a statistic?

Solutions:

The sample is the 500 shoppers questioned. The population is less clear. While the intended population of this survey was Tacoma citizens, the effective population was mall shoppers. There is no reason to assume that the 500 shoppers questioned would be representative of all Tacoma citizens.

The 60% value was based on the sample, so it is a statistic.

The examples on this page are detailed in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=143

Try It

To determine the average length of trout in a lake, researchers catch 20 fish and measure them. What is the sample and population in this study?

Solution:

The sample is the 20 fish caught. The population is all fish in the lake. The sample may be somewhat unrepresentative of the population since not all fish may be large enough to catch the bait.

A college reports that the average age of their students is 28 years old. Is this a statistic or a parameter?

Solution:

This is a parameter, since the college would have access to data on all students (the population)

Categorizing Data

Quantitative or Categorical

Once we have gathered data, we might wish to classify it. Roughly speaking, data can be classified as categorical data or quantitative data.

Quantitative and categorical data

Categorical (qualitative) data are pieces of information that allow us to classify the objects under investigation into various categories.

Quantitative data are responses that are numerical in nature and with which we can perform meaningful arithmetic calculations.

example

We might conduct a survey to determine the name of the favorite movie that each person in a math class saw in a movie theater.

When we conduct such a survey, the responses would look like: *Finding Nemo*, *The Hulk*, or *Terminator 3: Rise of the Machines*. We might count the number of people who give each answer, but the answers themselves do not have any numerical values: we cannot perform computations with an answer like "*Finding Nemo*." Is this categorical or quantitative data?

Solution:

This would be categorical data.

Example

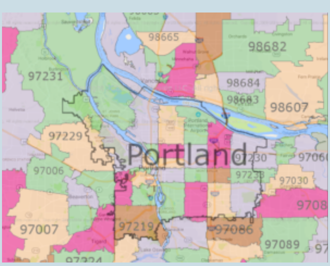
A survey could ask the number of movies you have seen in a movie theater in the past 12 months (0, 1, 2, 3, 4, . . .). Is this categorical or quantitative data?

Solution:

This would be quantitative data.Other examples of quantitative data would be the running time of the movie you saw most recently (104 minutes, 137 minutes, 104 minutes, . . .) or the amount of money you paid for a movie ticket the last time you went to a movie theater ($5.50, $7.75, $9, . . .).

Sometimes, determining whether or not data is categorical or quantitative can be a bit trickier. In the next example, teh data collected is in numerical form, but it is not quantitative data. Read on to find out why.

example



Zip Codes for Portland, OR

Suppose we gather respondents' ZIP codes in a survey to track their geographical location. Is this categorical or quantitative?

Solution:

ZIP codes are numbers, but we can't do any meaningful mathematical calculations with them (it doesn't make sense to say that 98036 is "twice" 49018 — that's like saying that Lynnwood, WA is "twice" Battle Creek, MI, which doesn't make sense at all), so ZIP codes are really categorical data.

Example

A survey about the movie you most recently attended includes the question "How would you rate the movie you just saw?" with these possible answers:

1 – it was awful

2 – it was just OK

3 – I liked it

4 – it was great

5 – best movie ever!

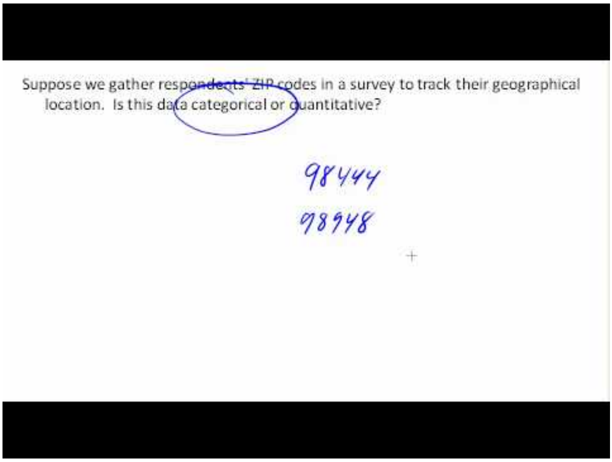
Is this categorical or quantitative?

Solution:

Again, there are numbers associated with the responses, but we can't really do any calculations with them: a movie that rates a 4 is not necessarily twice as good as a movie that rates a 2, whatever that means; if two people see the movie and one of them thinks it stinks and the other thinks it's the best ever it doesn't necessarily make sense to say that "on average they liked it."

As we study movie-going habits and preferences, we shouldn't forget to specify the population under consideration. If we survey 3-7 year-olds the runaway favorite might be *Finding Nemo*. 13-17 year-olds might prefer *Terminator 3*. And 33-37 year-olds might prefer . . . well, *Finding Nemo*.

The examples in this page are discussed further in the following video:



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=143

Try It

Classify each measurement as categorical or quantitative.

1. Eye color of a group of people
2. Daily high temperature of a city over several weeks

3. Annual income

Solutions:

1. Categorical. 2. Quantitative 3. Quantitative

Attributions

This chapter contains material taken from *Math in Society* (on OpenTextBookStore) by David Lippman, and is used under a CC Attribution-Share Alike 3.0 United States (CC BY-SA 3.0 US) license.

This chapter contains material taken from of *Math for the Liberal Arts* (on Lumen Learning) by Lumen Learning, and is used under a *CC BY: Attribution* license.

SAMPLING AND EXPERIMENTATION

Learning Outcomes

- Identify methods for obtaining a random sample of the intended population of a study
- Identify ineffective ways of obtaining a random sample from a population
- Identify types of sample bias
- Identify the differences between observational study and an experiment
- Identify the treatment in an experiment
- Determine whether an experiment may have been influenced by confounding

As we mentioned previously, the first thing we should do before conducting a survey is to identify the population that we want to study. In this lesson, we will show you examples of how to identify the population in a study, and determine whether or not the study actually represents the intended population. We will discuss different techniques for *random sampling* that are intended to ensure a population is well represented in a sample.

We will also identify the difference between an observational study and an experiment, and ways experiments can be conducted. By the end of this lesson, we hope that you will also be confident in identifying when an experiment may have been affected by confounding or the placebo effect, and the methods that are employed to avoid them.

Sampling Methods and Bias

Selecting a Population

Suppose we are hired by a politician to determine the amount of support he has among the electorate should he decide to run for another term. What population should we study? Every person in the district? Not every person is eligible to vote, and regardless of how strongly someone likes or dislikes the candidate, they don't have much to do with him being re-elected if they are not able to vote.

What about eligible voters in the district? That might be better, but if someone is eligible to vote but does not register by the deadline, they won't have any say in the election either. What about registered voters? Many people are registered but choose not to vote. What about "likely voters?"

This is the criteria used in much political polling, but it is sometimes difficult to define a "likely voter." Is it someone who voted in the last election? In the last general election? In the last presidential election? Should we consider someone who just turned 18 a "likely voter?" They weren't eligible to vote in the past, so how do we judge the likelihood that they will vote in the next election?

In November 1998, former professional wrestler Jesse "The Body" Ventura was elected governor of Minnesota. Up until right before the election, most polls showed he had little chance of winning. There were several contributing factors to the polls not reflecting the actual intent of the electorate:

- Ventura was running on a third-party ticket and most polling methods are better suited to a two-candidate race.

- Many respondents to polls may have been embarrassed to tell pollsters that they were planning to vote for a professional wrestler.
- The mere fact that the polls showed Ventura had little chance of winning might have prompted some people to vote for him in protest to send a message to the major-party candidates.

But one of the major contributing factors was that Ventura recruited a substantial amount of support from young people, particularly college students, who had never voted before and who registered specifically to vote in the gubernatorial election. The polls did not deem these young people likely voters (since in most cases young people have a lower rate of voter registration and a turnout rate for elections) and so the polling samples were subject to **sampling bias**: they omitted a portion of the electorate that was weighted in favor of the winning candidate.

Sampling bias

A sampling method is biased if every member of the population doesn't have equal likelihood of being in the sample.

So even identifying the population can be a difficult job, but once we have identified the population, how do we choose an appropriate sample? Remember, although we would prefer to survey all members of the population, this is usually impractical unless the population is very small, so we choose a sample. There are many ways to sample a population, but there is one goal we need to keep in mind: we would like the sample to be *representative of the population*.

Returning to our hypothetical job as a political pollster, we would not anticipate very accurate results if we drew all of our samples from among the customers at a Starbucks, nor would we expect that a sample drawn entirely from the membership list of the local Elks club would provide a useful picture of district-wide support for our candidate.

One way to ensure that the sample has a reasonable chance of mirroring the population is to employ *randomness*. The most basic random method is simple random sampling.

Simple random sample

A **random sample** is one in which each member of the population has an equal probability of being chosen. A **simple random sample** is one in which every member of the population and any group of members has an equal probability of being chosen.

example

If we could somehow identify all likely voters in the state, put each of their names on a piece of paper, toss the slips into a (very large) hat and draw 1000 slips out of the hat, we would have a simple random sample.

In practice, computers are better suited for this sort of endeavor than millions of slips of paper and extremely large headgear.

It is always possible, however, that even a random sample might end up not being totally representative of the population. If we repeatedly take samples of 1000 people from among the population of likely voters in the state of Washington, some of these samples might tend to have a slightly higher percentage of Democrats (or Republicans) than does the general population; some samples might include more older people and some samples might include more younger people; etc. In most cases, this **sampling variability** is not significant.

Sampling variability

The natural variation of samples is called **sampling variability**.

This is unavoidable and expected in random sampling, and in most cases is not an issue.

To help account for variability, pollsters might instead use a **stratified sample**.

Stratified sampling

In **stratified sampling**, a population is divided into a number of subgroups (or strata). Random samples are then taken from each subgroup with sample sizes proportional to the size of the subgroup in the population.

example

Suppose in a particular state that previous data indicated that the electorate was comprised of 39% Democrats, 37% Republicans and 24% independents. In a sample of 1000 people, they would then expect to get about 390 Democrats, 370 Republicans and 240 independents. To accomplish this, they could randomly select 390 people from among those voters known to be Democrats, 370 from those known to be Republicans, and 240 from those with no party affiliation.

Stratified sampling can also be used to select a sample with people in desired age groups, a specified mix ratio of males and females, etc. A variation on this technique is called **quota sampling**.

Quota sampling

Quota sampling is a variation on stratified sampling, wherein samples are collected in each subgroup until the desired quota is met.

example

Suppose the pollsters call people at random, but once they have met their quota of 390 Democrats, they only gather people who do not identify themselves as a Democrat.

You may have had the experience of being called by a telephone pollster who started by asking you your age, income, etc. and then thanked you for your time and hung up before asking any “real” questions. Most likely, they already had contacted enough people in your demographic group and were looking for people who were older or younger, richer or poorer, etc. Quota sampling is usually a bit easier than stratified sampling, but also does not ensure the same level of randomness.

Another sampling method is **cluster sampling**, in which the population is divided into groups, and one or more groups are randomly selected to be in the sample.

Cluster sampling

In **cluster sampling**, the population is divided into subgroups (clusters), and a set of subgroups are selected to be in the sample.

example

If the college wanted to survey students, since students are already divided into classes, they could randomly select 10 classes and give the survey to all the students in those classes. This would be cluster sampling.

Other sampling methods include **systematic sampling**.

Systematic sampling

In **systematic sampling**, every n^{th} member of the population is selected to be in the sample.

example

To select a sample using systematic sampling, a pollster calls every 100th name in the phone book.

Systematic sampling is not as random as a simple random sample (if your name is Albert Aardvark and your sister Alexis Aardvark is right after you in the phone book, there is no way you could both end up in the sample) but it can yield acceptable samples.

The Worst Way to Sample

Perhaps the worst types of sampling methods are **convenience samples** and **voluntary response samples**.

Convenience sampling and voluntary response sampling

Convenience sampling is the practice of samples chosen by selecting whoever is convenient.

Voluntary response sampling is allowing the sample to volunteer.

example

A pollster stands on a street corner and interviews the first 100 people who agree to speak to him. Which sampling method is represented by this scenario?

Solution:

This is a convenience sample.

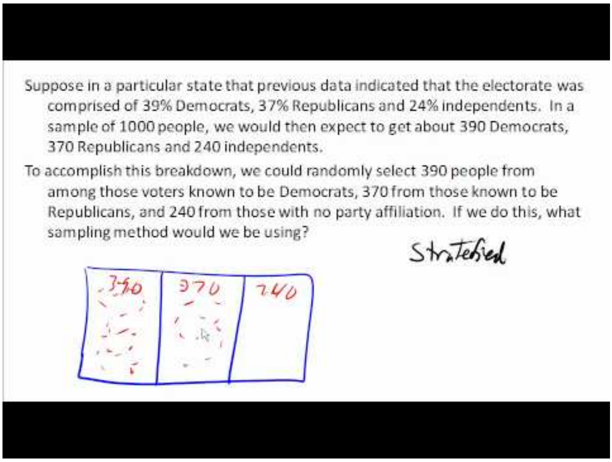
A website has a survey asking readers to give their opinion on a tax proposal. Which sampling method is represented?

Solution:

This is a self-selected sample, or voluntary response sample, in which respondents volunteer to participate.

Usually voluntary response samples are skewed towards people who have a particularly strong opinion about the subject of the survey or who just have way too much time on their hands and enjoy taking surveys.

Watch the following video for an overview of all the sampling methods discussed so far.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=145

Try It

[Click here to try this problem.]()

Problematic Sampling and Surveying

There are number of ways that a study can be ruined before you even start collecting data. The first we have already explored – **sampling** or **selection bias**, which is when the sample is not representative of the population. One example of this is **voluntary response bias**, which is bias introduced by only collecting data from those who volunteer to participate. This is not the only potential source of bias.

Sources of bias

- **Sampling bias** – when the sample is not representative of the population
- **Voluntary response bias** – the sampling bias that often occurs when the sample is volunteers
- **Self-interest study** – bias that can occur when the researchers have an interest in the outcome
- **Response bias** – when the responder gives inaccurate responses for any reason
- **Perceived lack of anonymity** – when the responder fears giving an honest answer might negatively affect them
- **Loaded questions** – when the question wording influences the responses
- **Non-response bias** – when people

refusing to participate in the study can influence the validity of the outcome

examples

Consider a recent study which found that chewing gum may raise math grades in teenagers[1]. This study was conducted by the Wrigley Science Institute, a branch of the Wrigley chewing gum company. Identify the type of sampling bias found in this example.

Solution:

This is an example of a **self-interest study**; one in which the researchers have a vested interest in the outcome of the study. While this does not necessarily ensure that the study was biased, it certainly

1. Reuters. http://news.yahoo.com/s/nm/20090423/od_uk_nm/oukoe_uk_gum_learning. Retrieved 4/27/09

suggests that we should subject the study to extra scrutiny.

A survey asks people "when was the last time you visited your doctor?" What type of sampling bias might this lead to?

Solution:

This might suffer from **response bias**, since many people might not remember exactly when they last saw a doctor and give inaccurate responses.

Sources of response bias may be innocent, such as bad memory, or as intentional as pressuring by the pollster. Consider, for example, how many voting initiative petitions people sign without even reading them.

A survey asks participants a question about their interactions with members of other races. Which sampling bias might occur for this survey strategy?

Solution:

Here, a **perceived lack of anonymity** could influence the outcome. The respondent might not want to be perceived as racist even if they are, and give an untruthful answer.

An employer puts out a survey asking their employees if they have a drug abuse problem and need treatment help. Which sampling bias may occur in this scenario?

Solution:

Here, answering truthfully might have consequences; responses might not be accurate if the employees do not feel their responses are anonymous or fear retribution from their employer. This survey has the potential for **perceived lack of anonymity**.

A survey asks "do you support funding research of alternative energy sources to reduce our reliance on high-polluting fossil fuels?" Which sampling bias may result from this survey?

Solution:

This is an example of a **loaded** or **leading question** – questions whose wording leads the respondent towards an answer.

Loaded questions can occur intentionally by pollsters with an agenda, or accidentally through poor question wording. Also a concern is **question order**, where the order of questions changes the results. A psychology researcher provides an example[2]:

> "My favorite finding is this: we did a study where we asked students, 'How satisfied are you with your life? How often do you have a date?' The two answers were not statistically related – you would conclude that there is no relationship between dating frequency and life satisfaction. But when we reversed the order and asked, 'How often do you have a date? How satisfied are you with your life?' the statistical relationship was a strong one. You would now conclude that there is nothing as important in a student's life as dating frequency."

---

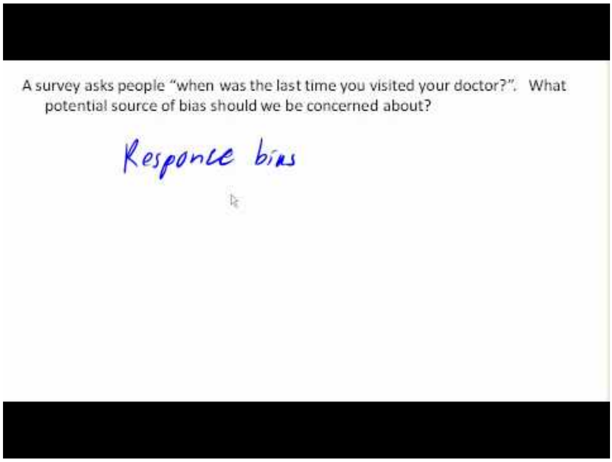
2. Swartz, Norbert. http://www.umich.edu/~newsinfo/MT/01/Fal01/mt6f01.html. Retrieved 3/31/2009

A telephone poll asks the question "Do you often have time to relax and read a book?", and 50% of the people called refused to answer the survey. Which sampling bias is represented by this survey?

Solution:

It is unlikely that the results will be representative of the entire population. This is an example of **non-response bias**, introduced by people refusing to participate in a study or dropping out of an experiment. When people refuse to participate, we can no longer be so certain that our sample is representative of the population.

These problematic scenarios for statistics gathering are discussed further in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=145

Try It

[Click here to try this problem.]()

Experiments

Observing vs. Acting

So far, we have primarily discussed **observational studies** – studies in which conclusions would be drawn from observations of a sample or the population. In some cases these observations might be unsolicited, such as studying the percentage of cars that turn right at a red light even when there is a "no turn on red" sign. In other cases the observations are solicited, like in a survey or a poll.

In contrast, it is common to use **experiments** when exploring how subjects react to an outside influence. In an experiment, some kind of **treatment** is applied to the subjects and the results are measured and recorded.

Observational studies and experiments

- An **observational study** is a study based on observations or measurements
- An **experiment** is a study in which the effects of a **treatment** are measured

Examples

Here are some examples of experiments:

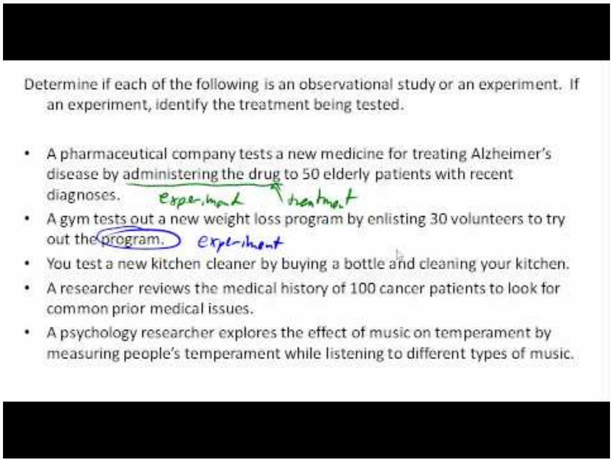
A pharmaceutical company tests a new medicine for treating Alzheimer's disease by administering the drug to 50 elderly patients with recent diagnoses. The treatment here is the new drug.

A gym tests out a new weight loss program by enlisting 30 volunteers to try out the program. The treatment here is the new program.

You test a new kitchen cleaner by buying a bottle and cleaning your kitchen. The new cleaner is the treatment.

A psychology researcher explores the effect of music on temperament by measuring people's temperament while listening to different types of music. The music is the treatment.

These examples are discussed further in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=145

Try It

Is each scenario describing an observational study or an experiment?

a. The weights of 30 randomly selected people are measured

b. Subjects are asked to do 20 jumping jacks, and then their heart rates are measured

c. Twenty coffee drinkers and twenty tea drinkers are given a concentration test

Solutions:

a. Observational study

b. Experiment; the treatment is the jumping jacks

c. Experiment; the treatments are coffee and tea

When conducting experiments, it is essential to isolate the treatment being tested.

example

Suppose a middle school (junior high) finds that their students are not scoring well on the state's standardized math test. They decide to run an experiment to see if an alternate curriculum would improve scores. To run the test, they hire a math specialist to come in and teach a class using the new curriculum. To their delight, they see an improvement in test scores.

The difficulty with this scenario is that it is not clear whether the curriculum is responsible for the improvement, or whether the improvement is due to a math specialist teaching the class. This is called **confounding** – when it is not clear which factor or factors caused the observed effect. Confounding is the downfall of many experiments, though sometimes it is hidden.

Confounding

Confounding occurs when there are two potential variables that could have caused the outcome and it is not possible to determine which actually caused the result.

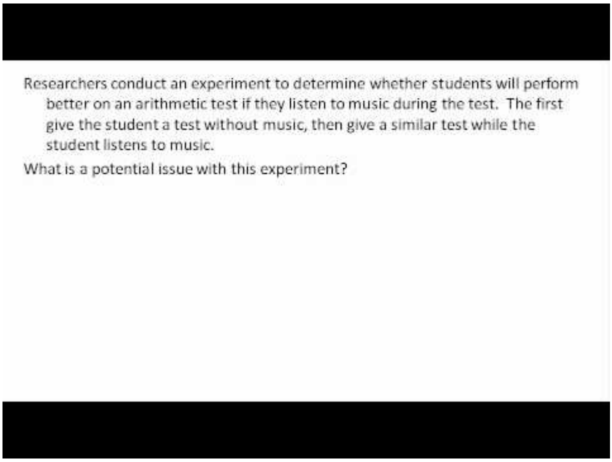
examples

A drug company study about a weight loss pill might report that people lost an average of 8 pounds while using their new drug. However, in the fine print you find a statement saying that participants were encouraged to also diet and exercise. It is not clear in this case whether the weight loss is due to the pill, to diet and exercise, or a combination of both. In this case confounding has occurred.

Example

Researchers conduct an experiment to determine whether students will perform better on an arithmetic test if they listen to music during the test. They first give the student a test without music, then give a similar test while the student listens to music. In this case, the student might perform better on the second test, regardless of the music, simply because it was the second test and they were warmed up.

View the following for additional discussion of these examples.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=145

Try It

Click here to try this problem.

There are a number of measures that can be introduced to help reduce the likelihood of confounding. The primary measure is to use a **control group**.

Control group

When using a control group, the participants are divided into two or more groups, typically a **control group** and a treatment group. The treatment group receives the treatment being tested; the control group does not receive the treatment.

Ideally, the groups are otherwise as similar as possible, isolating the treatment as the only potential source of difference between the groups. For this reason, the method of dividing groups is important. Some researchers attempt to ensure that the groups have similar characteristics (same number of females, same number of people over 50, etc.), but it is nearly impossible to control for every characteristic. Because of this, random assignment is very commonly used.

examples

To determine if a two day prep course would help high school students improve their scores on the SAT test, a group of students was randomly divided into two subgroups. The first group, the treatment group, was given a two day prep course. The second group, the control group, was not given the prep course. Afterwards, both groups were given the SAT.

A company testing a new plant food grows two crops of plants in adjacent fields, the treatment group receiving the new plant food and the control group not. The crop yield would then be compared. By growing them at the same time in adjacent fields, they are controlling for weather and other confounding factors.

Sometimes not giving the control group anything does not completely control for confounding variables. For example, suppose a medicine study is testing a new headache pill by giving the treatment group the pill and the control group nothing. If the treatment group showed improvement, we would not know whether it was due to the medicine in the pill, or a response to have taken any pill. This is called a **placebo effect**.

Placebo effect

The **placebo effect** is when the effectiveness of a treatment is influenced by the patient's perception of how effective they think the treatment will be, so a result might be seen even if the treatment is ineffectual.

example

A study found that when doing painful dental tooth extractions, patients told they were receiving a strong painkiller while actually receiving a saltwater injection found as much pain relief as patients receiving a dose of morphine.[3]

To control for the placebo effect, a **placebo**, or dummy treatment, is often given to the control group. This way, both groups are truly identical except for the specific treatment given.

Placebo and Placebo controlled experiments

- A **placebo** is a dummy treatment given to control for the placebo effect.
- An experiment that gives the control group a placebo is called a **placebo controlled experiment**.

3. Levine JD, Gordon NC, Smith R, Fields HL. (1981) Analgesic responses to morphine and placebo in individuals with postoperative pain. Pain. 10:379-89.

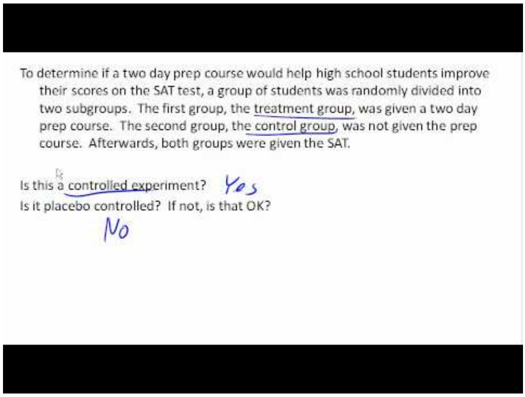
examples

In a study for a new medicine that is dispensed in a pill form, a sugar pill could be used as a placebo.

In a study on the effect of alcohol on memory, a non-alcoholic beer might be given to the control group as a placebo.

In a study of a frozen meal diet plan, the treatment group would receive the diet food, and the control could be given standard frozen meals stripped of their original packaging.

The following video walks through the controlled experiment scenarios, including the ones using placebos.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=145

In some cases, it is more appropriate to compare to a conventional treatment than a placebo. For example, in a cancer research study, it would not be ethical to deny any treatment to the control group or to give a placebo treatment. In this case, the currently acceptable medicine would be given to the second group, called a **comparison group** in this case. In our SAT test example, the non-treatment group would most likely be encouraged to study on their own, rather than be asked to not study at all, to provide a meaningful comparison.

When using a placebo, it would defeat the purpose if the participant knew they were receiving the placebo.

Blind studies

- A **blind study** is one in which the participant does not know whether or not they are receiving the treatment or a placebo.
- A **double-blind study** is one in which those interacting with the participants don't know who is in the treatment group and who is in the control group.

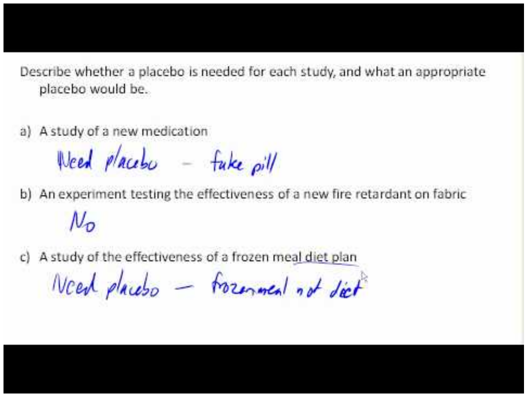
examples

In a study about anti-depression medicine, you would not want the psychological evaluator to know whether the patient is in the treatment or control group either, as it might influence their evaluation, so the experiment should be conducted as a double-blind study.

It should be noted that not every experiment needs a control group.

If a researcher is testing whether a new fabric can withstand fire, she simply needs to torch multiple samples of the fabric – there is no need for a control group.

These examples are demonstrated in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=145

Try It now

To test a new lie detector, two groups of subjects are given the new test. One group is asked to answer all the questions truthfully, and the second group is asked to lie on one set of questions. The person administering the lie detector test does not know what group each subject is in.

Does this experiment have a control group? Is it blind, double-blind, or neither?

Solution:

The truth-telling group could be considered the control group, but really both groups are treatment groups here, since it is important for the lie detector to be able to correctly identify lies, and also not identify truth telling as lying. This study is blind, since the person running the test does not know what group each subject is in.

Attributions

PUTTING IT TOGETHER: STATISTICS: COLLECTING DATA

You've been learning about ways to make sense of the overload of data that constantly surrounds you. Now that you've completed the module, you can better determine how to collect data and determine what the data truly represent.

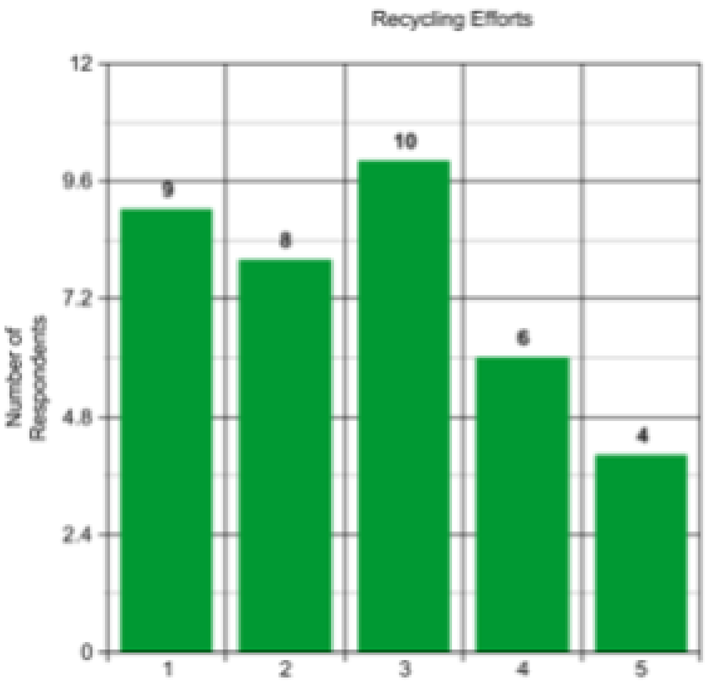
At some point, you have probably been involved in some type of attempt to collect data. Perhaps you were asked to take a "quick" survey about your attitude toward some type of political or consumer issue, and you agreed.

 The survey took longer than promised, but at least you helped the person collect random and unbiased data. Or did you? In order to appreciate the data collected, you need to ask yourself a few questions:

- Is it possible to get a truly random sample from a phone survey?
- What is the population of the sample?
- What are possible sources of bias?
- Is the data being collected from you a statistic or a parameter? Is it categorical or qualitative?
- Is this an experiment or observational study?

Calling a list of phone numbers is partly random, but there is potential to leave out part of the population. For example, a phone survey misses people who don't have a landline and use only a cell phone. It also misses people who monitor calls and let the answering machine pick up. So the sample you were part of wasn't truly random.

Let's consider some of the results, which are mailed to you at a later date. The first thing you see is this bar graph. What does it suggest?



1= never, 5= as much as possible

When you look closely, you notice that the data does not indicate a specific quantitative amount, like how many times a week the respondent recycles. The survey instead takes qualitative data – self-reflection about likeliness to recycle – and quantifies that information on a scale of 1 to 5.

You can also determine that the majority of respondents think they recycle less than, or equal to average. It also shows that the category selected by the most survey-takers is 3, halfway between recycling as much as possible and never recycling. But is this what the general public truly thinks about recycling?

To find out, you need to take this research into your hands by considering additional methods of data collection. If you are interested in qualitative data, you must be prepared for a little more time-consuming research. The main methods of conducting this research involve individual interviews, focus groups, and direct observation. Maybe you might ask people at random around town. Or you might prefer quantitative data, and count the number of recycle bins placed outside homes on the proper day. But be careful not to bias your results. If you include data collected in a town that does not have recycling pickup, you will most likely obtain different results than in an area where recycling pickup is easy and free.

Collecting unbiased, useful data is a challenging task. You must always be sure to take possible errors into account and design your data collection method to minimize them. How would you design a method to collect data about society's dedication to recycling?

Attributions

EXERCISES

Exercises:

Skills

1. A political scientist surveys 28 of the current 106 representatives in a state's congress. Of them, 14 said they were supporting a new education bill, 12 said there were not supporting the bill, and 2 were undecided.
   a. What is the population of this survey?
   b. What is the size of the population?
   c. What is the size of the sample?
   d. Give the sample statistic for the proportion of voters surveyed who said they were supporting the education bill.
   e. Based on this sample, we might expect how many of the representatives to support the education bill?
2. The city of Raleigh has 9500 registered voters. There are two candidates for city council in an upcoming election: Brown and Feliz. The day before the election, a telephone poll of 350 randomly selected registered voters was conducted. 112 said they'd vote for Brown, 207 said they'd vote for Feliz, and 31 were undecided.

   a. What is the population of this survey?
   b. What is the size of the population?
   c. What is the size of the sample?
   d. Give the sample statistic for the proportion of voters surveyed who said they'd vote for Brown.
   e. Based on this sample, we might expect how many of the 9500 voters to vote for Brown?
3. Identify the most relevant source of bias in this situation: A survey asks the following: Should the mall prohibit loud and annoying rock music in clothing stores catering to teenagers?
4. Identify the most relevant source of bias in this situation: To determine opinions on voter support for a downtown renovation project, a surveyor randomly questions people working in downtown businesses.
5. Identify the most relevant source of bias in this situation: A survey asks people to report their actual income and the income they reported on their IRS tax form.
6. Identify the most relevant source of bias in this situation: A survey randomly calls people from the phone book and asks them to answer a long series of questions.
7. Identify the most relevant source of bias in this situation: A survey asks the following: Should the death penalty be permitted if innocent people might die?
8. Identify the most relevant source of bias in this

situation: A study seeks to investigate whether a new pain medication is safe to market to the public. They test by randomly selecting 300 men from a set of volunteers.

9. In a study, you ask the subjects their age in years. Is this data qualitative or quantitative?
10. In a study, you ask the subjects their gender. Is this data qualitative or quantitative?
11. Does this describe an observational study or an experiment: The temperature on randomly selected days throughout the year was measured.
12. Does this describe an observational study or an experiment? A group of students are told to listen to music while taking a test and their results are compared to a group not listening to music.
13. In a study, the sample is chosen by separating all cars by size, and selecting 10 of each size grouping. What is the sampling method?
14. In a study, the sample is chosen by writing everyone's name on a playing card, shuffling the deck, then choosing the top 20 cards. What is the sampling method?
15. A team of researchers is testing the effectiveness of a new HPV vaccine. They randomly divide the subjects into two groups. Group 1 receives new HPV vaccine, and Group 2 receives the existing HPV vaccine. The patients in the study do not know which group they are in.
    a. Which is the treatment group?

   b. Which is the control group (if there is one)?
   c. Is this study blind, double-blind, or neither?
   d. Is this best described as an experiment, a controlled experiment, or a placebo controlled experiment?

16. For the clinical trials of a weight loss drug containing *Garcinia cambogia t*he subjects were randomly divided into two groups. The first received an inert pill along with an exercise and diet plan, while the second received the test medicine along with the same exercise and diet plan. The patients do not know which group they are in, nor do the fitness and nutrition advisors.
   a. Which is the treatment group?
   b. Which is the control group (if there is one)?
   c. Is this study blind, double-blind, or neither?
   d. Is this best described as an experiment, a controlled experiment, or a placebo controlled experiment?

Concepts

17. A teacher wishes to know whether the males in his/her class have more conservative attitudes than the females. A questionnaire is distributed assessing attitudes.
   a. Is this a sampling or a census?
   b. Is this an observational study or an experiment?
   c. Are there any possible sources of bias in this study?
18. A study is conducted to determine whether people learn better with spaced or massed practice. Subjects volunteer

from an introductory psychology class. At the beginning of the semester 12 subjects volunteer and are assigned to the massed-practice group. At the end of the semester 12 subjects volunteer and are assigned to the spaced-practice condition.

   a. Is this a sampling or a census?
   b. Is this an observational study or an experiment?
   c. This study involves two kinds of non-random sampling: (1) Subjects are not randomly sampled from some specified population and (2) Subjects are not randomly assigned to groups. Which problem is more serious? What affect on the results does each have?

19. A farmer believes that playing Barry Manilow songs to his peas will increase their yield. Describe a controlled experiment the farmer could use to test his theory.
20. A sports psychologist believes that people are more likely to be extroverted as adults if they played team sports as children. Describe two possible studies to test this theory. Design one as an observational study and the other as an experiment. Which is more practical?

Exploration

21. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently

under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average length of time in months patients live once starting the treatment. Two researchers each follow a different set of 50 AIDS patients from the start of treatment until their deaths.

a. What is the population of this study?
b. List two reasons why the data may differ.
c. Can you tell if one researcher is correct and the other one is incorrect? Why?
d. Would you expect the data to be identical? Why or why not?
e. If the first researcher collected her data by randomly selecting 40 states, then selecting 1 person from each of those states. What sampling method is that?
f. If the second researcher collected his data by choosing 40 patients he knew. What sampling method would that researcher have used? What concerns would you have about this data set, based upon the data collection method?

22. Find a newspaper or magazine article, or the online equivalent, describing the results of a recent study (the results of a poll are not sufficient). Give a summary of the study's findings, then analyze whether the article provided enough information to determine the validity of the conclusions. If not, produce a list of things that

are missing from the article that would help you determine the validity of the study. Look for the things discussed in the text: population, sample, randomness, blind, control, placebos, etc.

Attributions

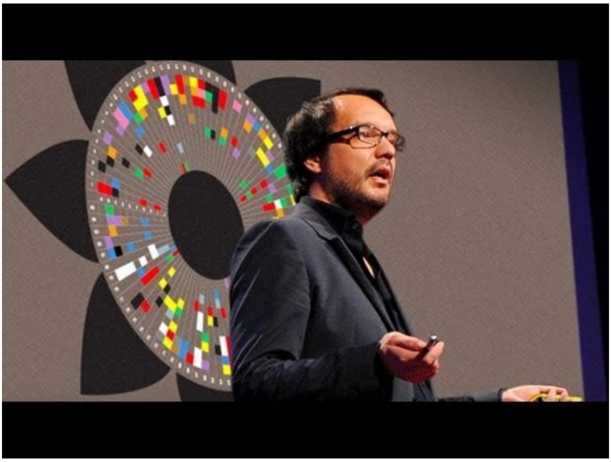
CHAPTER V

CHAPTER 5: STATISTICS: DESCRIBING DATA

WHY IT MATTERS: STATISTICS: DESCRIBING DATA

Why learn to present, analyze, and describe data?

The following video provides many examples of how contextualized, visually beautiful numbers add meaning and relevance to our life.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=151

In the following pages, we'll explore how to make information meaningful in a visual format. This will help us solve some problems more quickly, and allow us better understanding for how to compare different types of information to make better decisions as a result.

Attributions

REPRESENTING DATA GRAPHICALLY

Learning Outcomes

- Create a frequency table, bar graph, pareto chart, pictogram, or a pie chart to represent a data set
- Identify features of ineffective representations of data
- Create a histogram, pie chart, or frequency polygon that represents numerical data
- Create a graph that compares two quantities

In this lesson we will present some of the most common ways data is represented graphically. W e will also discuss some of the ways you can increase the accuracy and effectiveness of graphs of data that you create.

Presenting Categorical Data

Graphically

Visualizing Data

Categorical, or qualitative, data are pieces of information that allow us to classify the objects under investigation into various categories. We usually begin working with categorical data by summarizing the data into a **frequency table.**

Frequency Table

A frequency table is a table with two columns. One column lists the categories, and another for the frequencies with which the items in the categories occur (how many items fit into each category).

Example

An insurance company determines vehicle insurance premiums based on known risk factors. If a person is considered a higher risk, their premiums will be higher. One potential factor is the color of your car. The insurance company believes that people with some color cars are more likely to get in accidents. To research this, they examine police reports for recent total-loss collisions. The data is summarized in the frequency table below.

Color	Frequency
Blue	25
Green	52
Red	41
White	36
Black	39
Grey	23

Try It

Click here to try this problem.

Sometimes we need an even more intuitive way of displaying data. This is where charts and graphs come in. There are many, many ways of displaying data graphically, but we will concentrate on one very useful type of graph called a bar graph. In this section we will work with bar graphs that display categorical data; the next section will be devoted to bar graphs that display quantitative data.

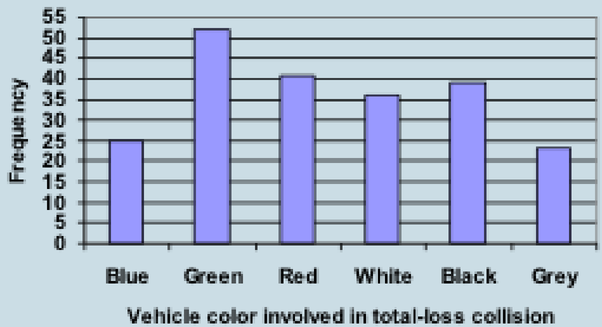
Bar graph

A **bar graph** is a graph that displays a bar for each category with the length of each bar indicating the frequency of that category.

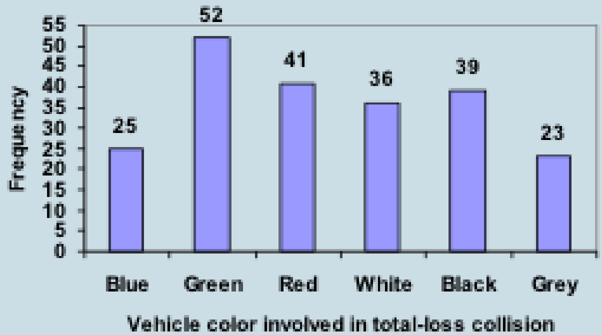
To construct a bar graph, we need to draw a vertical axis and a horizontal axis. The vertical direction will have a scale and measure the frequency of each category; the horizontal axis has no scale in this instance. The construction of a bar chart is most easily described by use of an example.

example

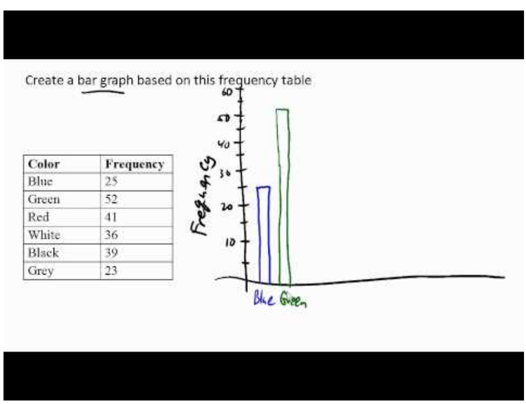
Using our car data from above, note the highest frequency is 52, so our vertical axis needs to go from 0 to 52, but we might as well use 0 to 55, so that we can put a hash mark every 5 units:



Notice that the height of each bar is determined by the frequency of the corresponding color. The horizontal gridlines are a nice touch, but not necessary. In practice, you will find it useful to draw bar graphs using graph paper, so the gridlines will already be in place, or using technology. Instead of gridlines, we might also list the frequencies at the top of each bar, like this:



The following video explains the process and value of moving data from a table to a bar graph.



A YouTube element has been excluded from this

version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

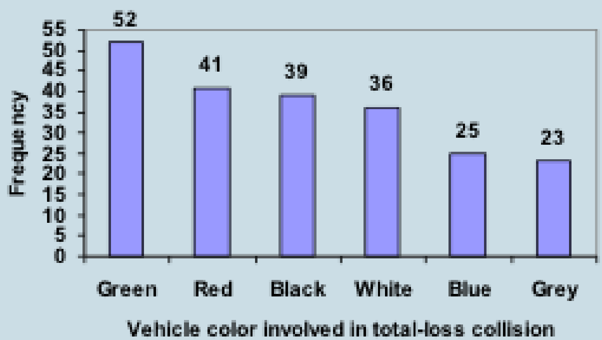
In this case, our chart might benefit from being reordered from largest to smallest frequency values. This arrangement can make it easier to compare similar values in the chart, even without gridlines. When we arrange the categories in decreasing frequency order like this, it is called a **Pareto chart**.

Pareto chart

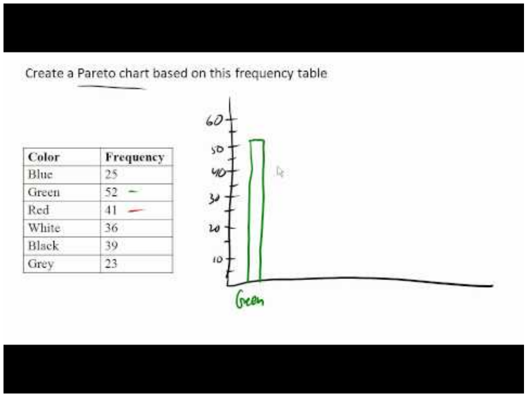
A **Pareto chart** is a bar graph ordered from highest to lowest frequency

Examples

Transforming our bar graph from earlier into a Pareto chart, we get:



The following video addressed Pareto charts.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

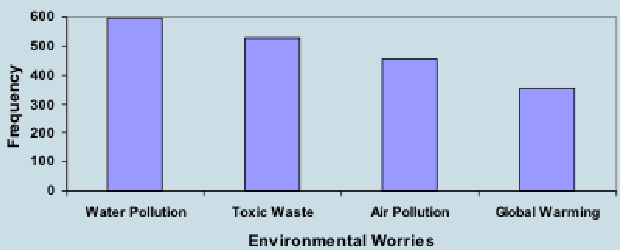
In a survey[1], adults were asked whether they personally worried about a variety of environmental concerns. The numbers (out of 1012 surveyed) who

1. Gallup Poll. March 5-8, 2009. http://www.pollingreport.com/enviro.htm

indicated that they worried "a great deal" about some selected concerns are summarized below.

Environmental Issue	Frequency
Pollution of drinking water	597
Contamination of soil and water by toxic waste	526
Air pollution	455
Global warming	354

This data could be shown graphically in a bar graph:



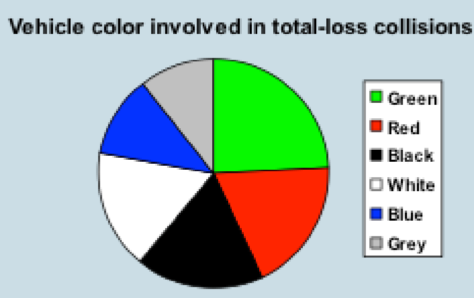
To show relative sizes, it is common to use a pie chart.

Pie Chart

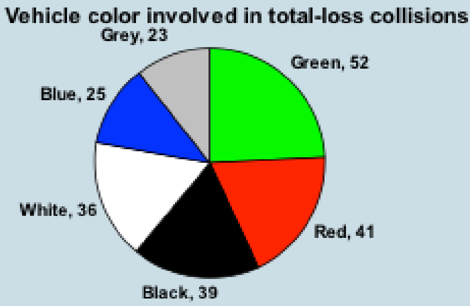
A **pie chart** is a circle with wedges cut of varying sizes marked out like slices of pie or pizza. The relative sizes of the wedges correspond to the relative frequencies of the categories.

examples

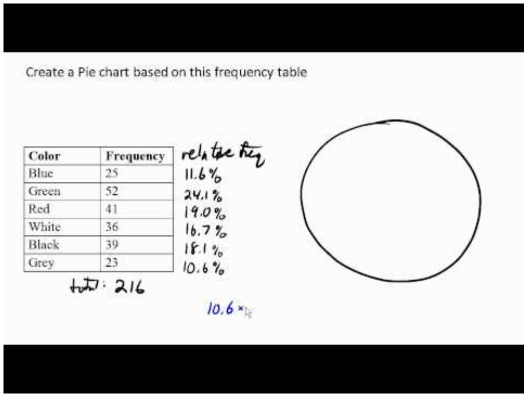
For our vehicle color data, a pie chart might look like this:



Pie charts can often benefit from including frequencies or relative frequencies (percents) in the chart next to the pie slices. Often having the category names next to the pie slices also makes the chart clearer.



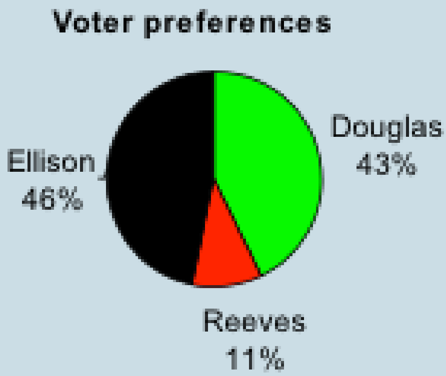
This video demonstrates how to **create** pie charts like the ones above.



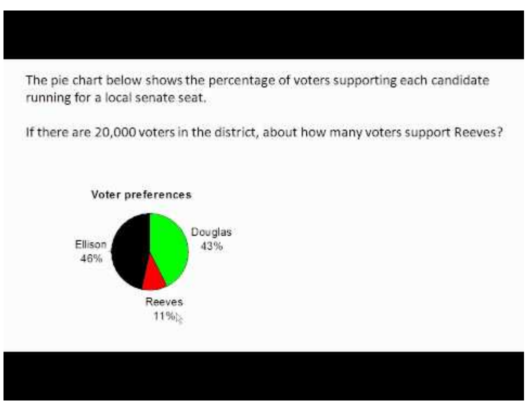
A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

The pie chart below shows the percentage of voters supporting each candidate running for a local senate seat.

If there are 20,000 voters in the district, the pie chart shows that about 11% of those, about 2,200 voters, support Reeves.



The following video addresses how to **read** a pie chart like the one above.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

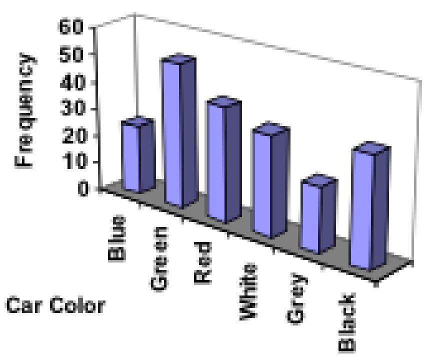
Pie charts look nice, but are harder to draw by hand than bar charts since to draw them accurately we would need to compute the angle each wedge cuts out of the circle, then measure the angle with a protractor. Computers are much better suited to drawing pie charts. Common software programs like Microsoft Word or Excel, OpenOffice.org Write or Calc, or Google Drive are able to create bar graphs, pie charts, and other graph types. It is suggested you refer back to the chapter on Excel and create your pie graphs using that software. If you are ever need to create a pie chart and do not have access to Excel, there are also numerous online tools besides Excel that can create graphs.[2].

2. For example: http://nces.ed.gov/nceskids/createAgraph/ or http://docs.google.com

Try It

Click here to try this problem.

Don't get fancy with graphs! People sometimes add features to graphs that don't help to convey their information. For example, 3-dimensional bar charts like the one shown below are usually not as effective as their two-dimensional counterparts.



Here is another way that fanciness can lead to trouble. Instead of plain bars, it is tempting to substitute meaningful images. This type of graph is called a **pictogram**.

Pictogram

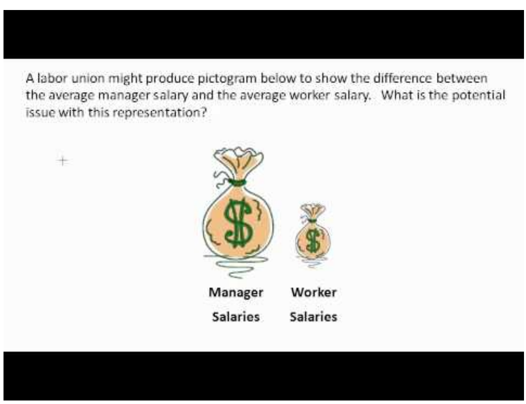
A **pictogram** is a statistical graphic in which the size of the picture is intended to represent the frequencies or size of the values being represented.

example

A labor union might produce the graph to the right to show the difference between the average manager salary and the average worker salary.

Looking at the picture, it would be reasonable to guess that the manager salaries is 4 times as large as the worker salaries – the area of the bag looks about 4 times as large. However, the manager salaries are in fact only twice as large as worker salaries, which were reflected in the picture by making the manager bag twice as tall.

This video reviews the two examples of ineffective data representation in more detail.



A YouTube element has been excluded from this

version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

Another distortion in bar charts results from setting the baseline to a value other than zero. The baseline is the bottom of the vertical axis, representing the least number of cases that could have occurred in a category. Normally, this number should be zero.

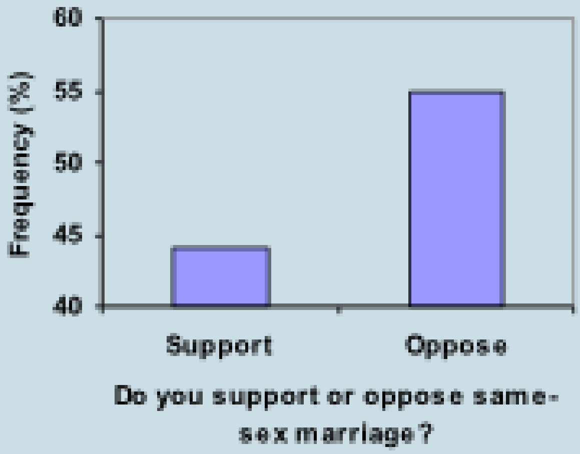
example

Compare the two graphs below showing support for same-sex marriage rights from a poll taken in December 2008[3]. The difference in the vertical scale on the first graph suggests a different story than

3. CNN/Opinion Research Corporation Poll. Dec 19-21, 2008, from http://www.pollingreport.com/civil.htm

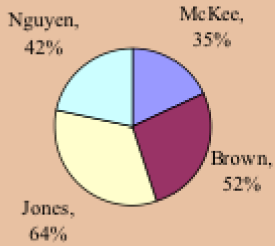
the true differences in percentages; the second graph makes it look like twice as many people oppose marriage rights as support it.


100
90
80
70
60
50
40
30
20
10
0
Frequency (%)
Support
Oppose
Do you support or oppose same-sex marriage?


60
55
50
45
40
Frequency (%)
Support
Oppose
Do you support or oppose same-sex marriage?

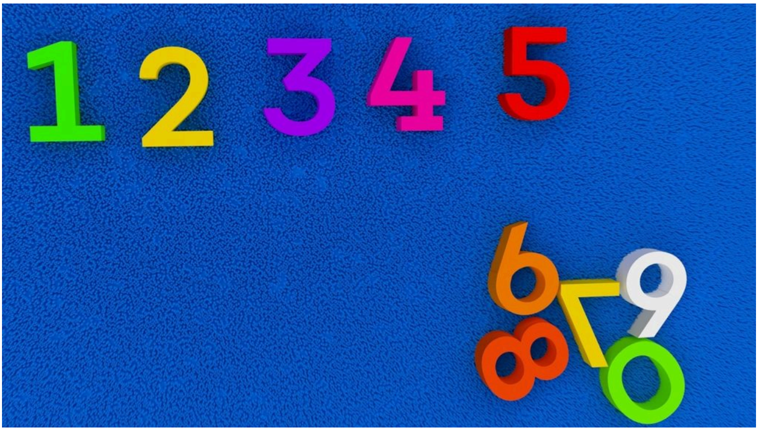
Try It

A poll was taken asking people if they agreed with the positions of the 4 candidates for a county office. Does the pie chart present a good representation of this data? Explain.



Presenting Quantitative Data Graphically

Visualizing Numbers



Quantitative, or numerical, data can also be summarized into frequency tables.

Example

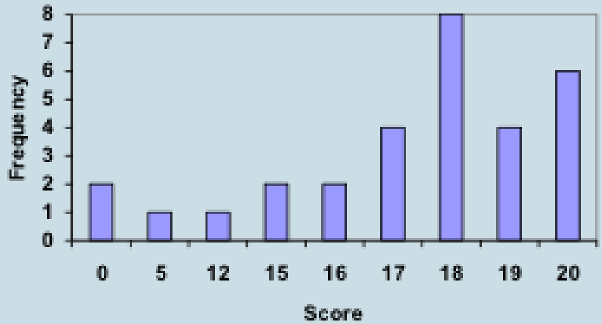
A teacher records scores on a 20-point quiz for the 30 students in his class. The scores are:

19 20 18 18 17 18 19 17 20 18 20 16 20 15 17 12 18 19 18 19 17 20 18 16 15 18 20 5 0 0

These scores could be summarized into a frequency table by grouping like values:

Score	Frequency
0	2
5	1
12	1
15	2
16	2
17	4
18	8
19	4
20	6

Using the table from the first example, it would be possible to create a standard bar chart from this summary, like we did for categorical data:



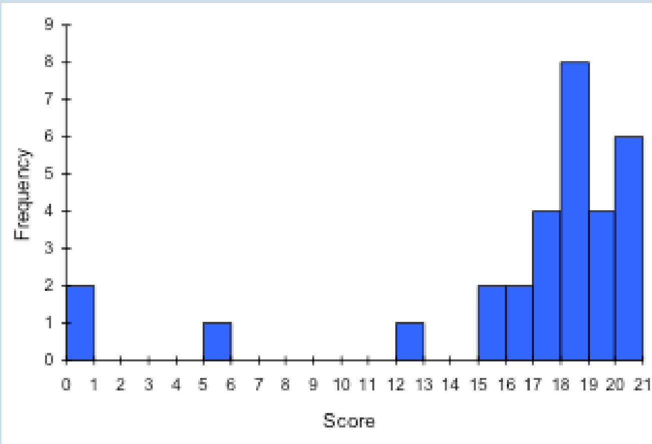
However, since the scores are numerical values, this chart doesn't really make sense; the first and second bars are five values apart, while the later bars are only one value apart. It would be more correct to treat the horizontal axis as a number line. This type of graph is called a **histogram**.

Histogram

A histogram is like a bar graph, but where the horizontal axis is a number line.

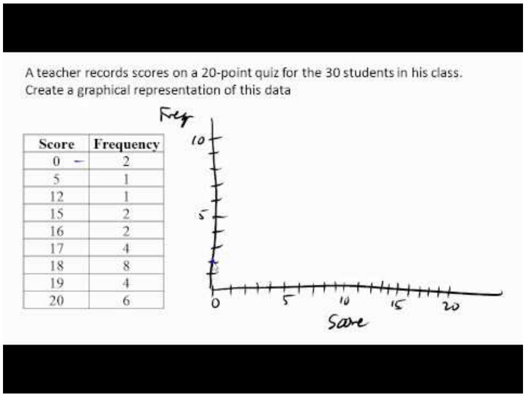
example

For the values above, a histogram would look like:



Notice that in the histogram, a bar represents values on the horizontal axis from that on the left hand-side of the bar up to, but not including, the value on the right hand side of the bar. Some people choose to have bars start at ½ values to avoid this ambiguity.

This video demonstrates the creation of the histogram by hand from this data.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

Fortunately, you can create a histogram using Excel. Refer back to the section on histograms in Chapter 2.

If we have a large number of widely varying data values, creating a frequency table that lists every possible value as a category would lead to an exceptionally long frequency table, and probably would not reveal any patterns. For this reason, it is common with quantitative data to group data into **class intervals**.

Class Intervals

Class intervals are groupings of the data. In general, we define class intervals so that

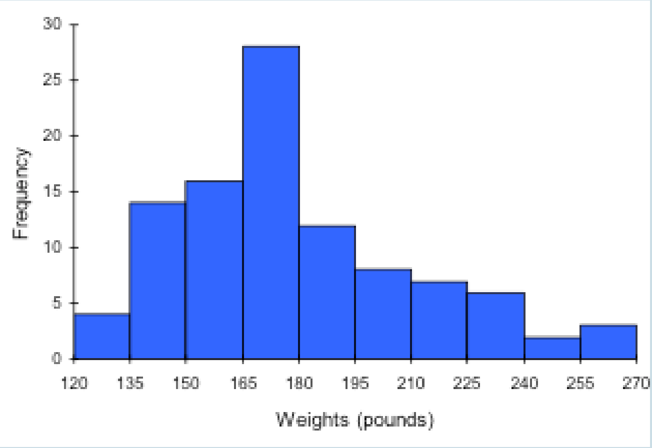
- each interval is equal in size. For example, if the first class contains values from 120-129, the second class should include values from 130-139.
- we have somewhere between 5 and 20 classes, typically, depending upon the number of data we're working with.

example

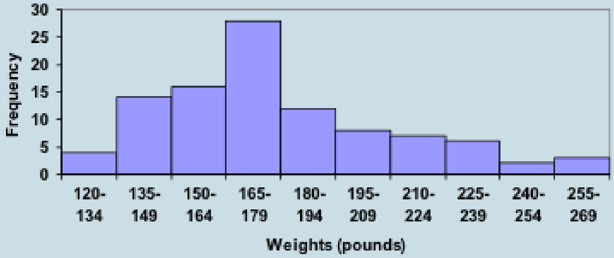
Suppose that we have collected weights from 100 male subjects as part of a nutrition study. For our weight data, we have values ranging from a low of 121 pounds to a high of 263 pounds, giving a total span of 263-121 = 142. We could create 7 intervals with a width of around 20, 14 intervals with a width of around 10, or somewhere in between. Often time we have to experiment with a few possibilities to find something that represents the data well. Let us try using an interval width of 15. We could start at 121, or at 120 since it is a nice round number.

Interval	Frequency
120 – 134	4
135 – 149	14
150 – 164	16
165 – 179	28
180 – 194	12
195 – 209	8
210 – 224	7
225 – 239	6
240 – 254	2
255 – 269	3

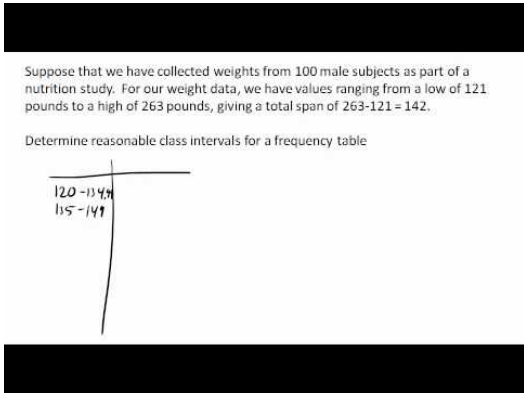
A histogram of this data would look like:



In many software packages, you can create a graph similar to a histogram by putting the class intervals as the labels on a bar chart.



The following video walks through this example in more detail.

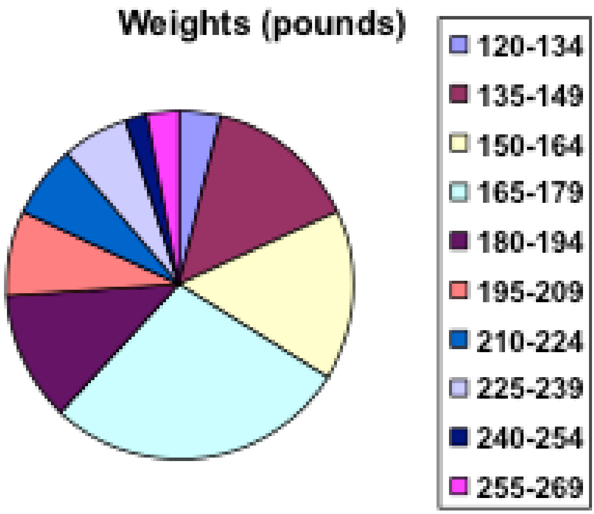


A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

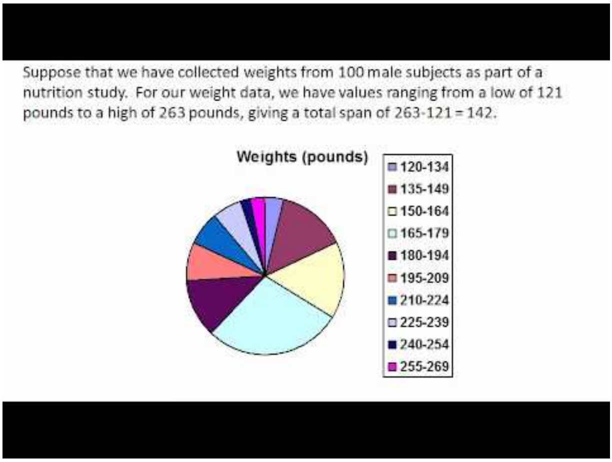
Try It

Click here to try this problem.

Other graph types such as pie charts are possible for quantitative data, but are not recommended. The usefulness of different graph types will vary depending upon the number of intervals and the type of data being represented. For example, a pie chart of our weight data is difficult to read because of the quantity of intervals we used.



To see more about why a pie chart isn't useful in this case, watch the following.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

Try It

The total cost of textbooks for the term was collected from 36 students. Create a histogram for this data.

$140 $160 $160 $165 $180 $220 $235
$240 $250 $260 $280 $285

$285 $285 $290 $300 $300 $305 $310
$310 $315 $315 $320 $320

$330 $340 $345 $350 $355 $360 $360
$380 $395 $420 $460 $460

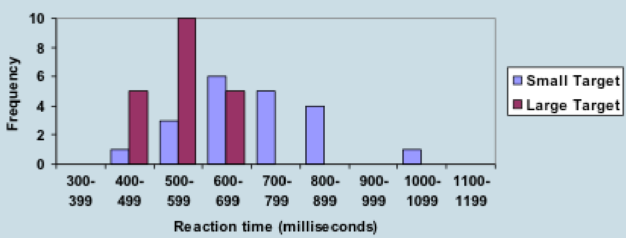
When collecting data to compare two groups, it is desirable to create a graph that compares quantities.

Example

The data below came from a task in which the goal is to move a computer mouse to a target on the screen as fast as possible. On 20 of the trials, the target was a small rectangle; on the other 20, the target was a large rectangle. Time to reach the target was recorded on each trial.

Interval (milliseconds)	Frequency small target	Frequency large target
300-399	0	0
400-499	1	5
500-599	3	10
600-699	6	5
700-799	5	0
800-899	4	0
900-999	0	0
1000-1099	1	0
1100-1199	0	0

One option to represent this data would be a comparative histogram or bar chart, in which bars for the small target group and large target group are placed next to each other.

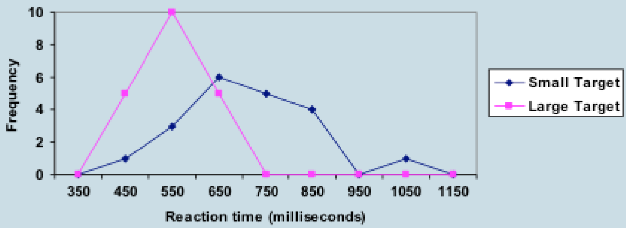


Frequency polygon

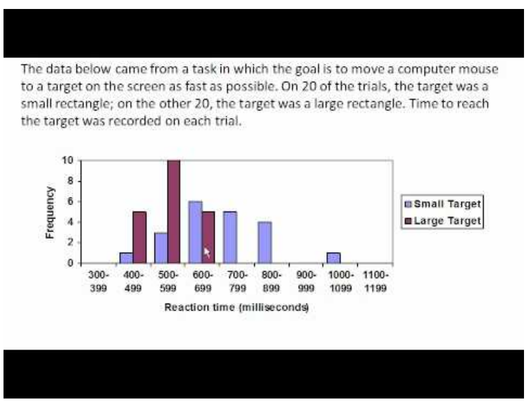
An alternative representation is a **frequency polygon**. A frequency polygon starts out like a histogram, but instead of drawing a bar, a point is placed in the midpoint of each interval at height equal to the frequency. Typically the points are connected with straight lines to emphasize the distribution of the data.

example

This graph makes it easier to see that reaction times were generally shorter for the larger target, and that the reaction times for the smaller target were more spread out.



The following video explains frequency polygon creation for this example.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=153

Attributions

NUMERICAL SUMMARIES OF DATA

Learning Outcomes

- Calculate the mean, median, and mode of a set of data
- Calculate the range of a data set, and recognize it's limitations in fully describing the behavior of a data set
- Calculate the standard deviation for a data set, and determine it's units
- Identify the difference between population variance and sample variance
- Identify the quartiles for a data set, and the calculations used to define them
- Identify the parts of a five number summary for a set of data, and create a box plot using it

It is often desirable to use a few numbers to summarize a data set. One important aspect of a set of data is where its center is located. In this lesson, measures of central tendency are discussed first. A second aspect of a distribution is how spread out it is. In other words, how much the data in the distribution vary from one another. The second section of this lesson describes measures of variability.

Measures of Central Tendency

Mean, Median, and Mode

Let's begin by trying to find the most "typical" value of a data set.

Note that we just used the word "typical" although in many cases you might think of using the word "average." We need to be careful with the word "average" as it means different things to different people in different contexts. One of the most common uses of the word "average" is what mathematicians and statisticians call the **arithmetic mean**, or just plain old **mean** for short. "Arithmetic mean" sounds rather fancy, but you have likely calculated a mean many times without realizing it; the mean is what most people think of when they use the word "average."

You learned how to compute the mean of a set of numbers in Excel. Refer back to Chapter 2 for instructions, if needed. It is perfectly okay to use Excel for these exercises. You can also use the Desmos calculator to compute a mean.

Mean

The **mean** of a set of data is the sum of the data values divided by the number of values.

examples

Marci's exam scores for her last math class were 79, 86, 82, and 94. What would the mean of these values would be?

Solution:

Typically we round means to one more decimal place than the original data had. In this case, we would round 85.25 to 85.3.

The number of touchdown (TD) passes thrown by each of the 31 teams in the National Football League in the 2000 season are shown below.

37 33 33 32 29 28 28 23 22 22 22 21 21 21 20

20 19 19 18 18 18 18 16 15 14 14 14 12 12 9 6

What is the mean number of TD passes?

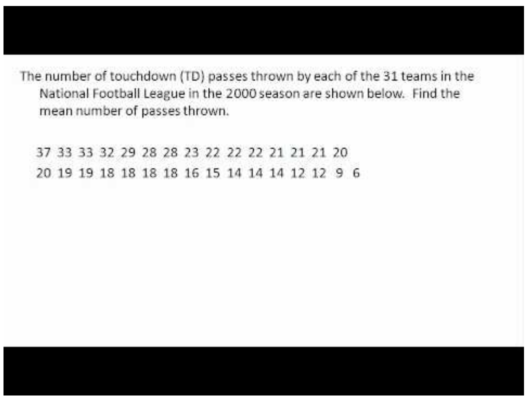
Solution:

Adding these values, we get 634 total TDs. Dividing by 31, the number of data values, we get 634/31 = 20.4516. It would be appropriate to round this to 20.5.

It would be most correct for us to report that "The mean number of touchdown passes thrown in the NFL in the 2000 season was 20.5 passes," but it is not uncommon to see the more casual word "average" used in place of "mean."

Remember that you can use Excel to compute the mean.

Both examples are described further in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

Try It

The price of a jar of peanut butter at 5 stores was \$3.29, \$3.59, \$3.79, \$3.75, and \$3.99. Find the mean price.

examples

The one hundred families in a particular neighborhood are asked their annual household income, to the nearest \$5 thousand dollars. The results are summarized in a frequency table below.

Income (thousands of dollars)	Frequency
15	6
20	8
25	11
30	17
35	19
40	20
45	12
50	7

What is the mean average income in this neighborhood?

Solution:

Calculating the mean by hand could get tricky if we try to type in all 100 values:

$$\frac{\overbrace{15+\cdots+15}^{6\text{terms}}+\overbrace{20+\cdots+20}^{8\text{terms}}+\overbrace{25+\cdots+25}^{11\text{terms}}+\ldots}{100}$$

We could calculate this more easily by noticing that adding 15 to itself six times is the same as = 90. Using this simplification, we get

$$\frac{15\cdot6+20\cdot8+25\cdot11+30\cdot17+35\cdot19+40\cdot20+45\cdot12+50\cdot7}{100}=\frac{3390}{100}=33.9$$

The mean household income of our sample is 33.9 thousand dollars ($33,900).

Extending off the last example, suppose a new family moves into the neighborhood example that has a household income of $5 million ($5000 thousand).

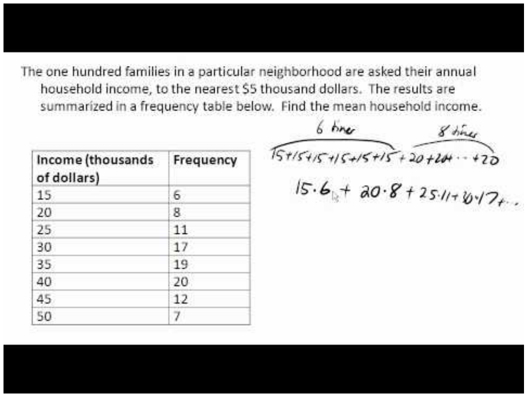
What is the new mean of this neighborhood's income?

Solution:

Adding this to our sample, our mean is now:

$$\frac{15\cdot 6+20\cdot 8+25\cdot 11+30\cdot 17+35\cdot 19+40\cdot 20+45\cdot 12+50\cdot 7+5000\cdot 1}{101}=\frac{8390}{101}=83.069$$

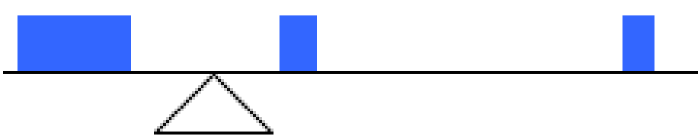
Both situations are explained further in this video.



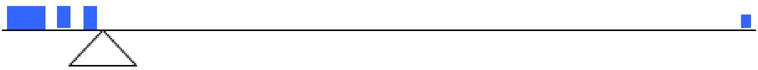
A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

While 83.1 thousand dollars ($83,069) is the correct mean household income, it no longer represents a "typical" value.

Imagine the data values on a see-saw or balance scale. The mean is the value that keeps the data in balance, like in the picture below.



If we graph our household data, the $5 million data value is so far out to the right that the mean has to adjust up to keep things in balance.



For this reason, when working with data that have **outliers** – values far outside the primary grouping – it is common to use a different measure of center, the **median**.

Median

The **median** of a set of data is the value in the middle when the data is in order.

- To find the median, begin by listing the data in order from smallest to largest, or largest to smallest.
- If the number of data values, *N*, is odd, then the median is the middle data value. This value can be found by rounding *N*/2 up to the next whole number.
- If the number of data values is even, there is no one middle value, so we find the

mean of the two middle values (values *N*/2 and *N*/2 + 1)

You may also use Excel to find the median of the data. Enter your data in a column, then use the command, =MEDIAN. You will need to select your data, just like you would do to find the mean of the data.

example

Returning to the football touchdown data, we would start by listing the data in order. Luckily, it was already in decreasing order, so we can work with it without needing to reorder it first.

37 33 33 32 29 28 28 23 22 22 22 21 21 21 20

20 19 19 18 18 18 18 16 15 14 14 14 12 12 9 6

What is the median TD value?

Solution:

Since there are 31 data values, an odd number, the median will be the middle number, the 16th data value (31/2 = 15.5, round up to 16, leaving 15 values below and 15 above). The 16th data value is 20, so the median number of touchdown passes in the 2000 season was 20 passes. Notice that for this data, the median is fairly close to the mean we calculated earlier, 20.5.[/hidden-answer]

Find the median of these quiz scores: 5 10 8 6 4 8 2 5 7 7

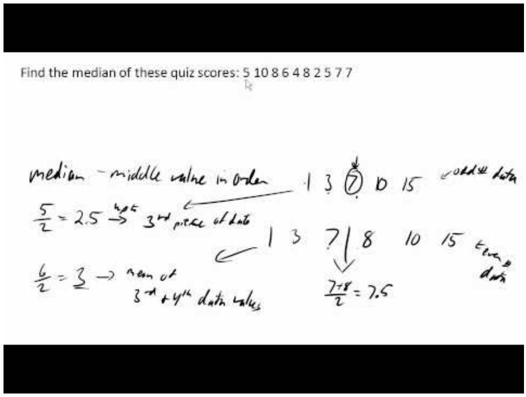
Solution:

We start by listing the data in order: 2 4 5 5 6 7 7 8 8 10

Since there are 10 data values, an even number, there is no one middle number. So we find the mean of the two middle numbers, 6 and 7, and get (6+7)/2 = 6.5.

The median quiz score was 6.5.

Learn more about these median examples in this video. Note that you may also use Excel to find the median in these problems.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

Try It

The price of a jar of peanut butter at 5 stores was \$3.29, \$3.59, \$3.79, \$3.75, and \$3.99. Find the median price.

Example

Let us return now to our original household income data

Income (thousands of dollars)	Frequency
15	6
20	8
25	11
30	17
35	19
40	20
45	12
50	7

What is the mean of this neighborhood's household income?

Solution:

Here we have 100 data values. If we didn't already know that, we could find it by adding the frequencies. Since 100 is an even number, we need to find the mean of the middle two data values – the

50th and 51st data values. To find these, we start counting up from the bottom:

There are 6 data values of $15, so Values 1 to 6 are $15 thousand

The next 8 data values are $20, so Values 7 to (6+8)=14 are $20 thousand

The next 11 data values are $25, so Values 15 to (14+11)=25 are $25 thousand

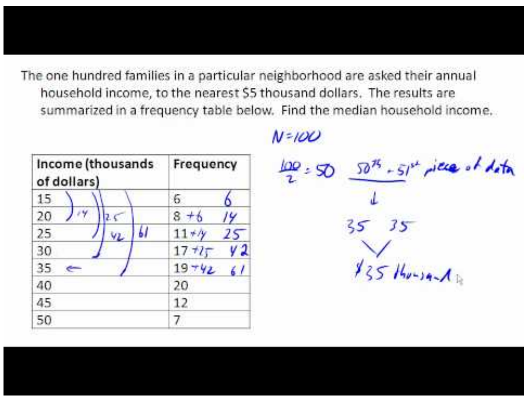
The next 17 data values are $30, so Values 26 to (25+17)=42 are $30 thousand

The next 19 data values are $35, so Values 43 to (42+19)=61 are $35 thousand

From this we can tell that values 50 and 51 will be $35 thousand, and the mean of these two values is $35 thousand. The median income in this neighborhood is $35 thousand.

If we add in the new neighbor with a $5 million household income, then there will be 101 data values, and the 51st value will be the median. As we discovered in the last example, the 51st value is $35 thousand. Notice that the new neighbor did not affect the median in this case. The median is not swayed as much by outliers as the mean is.

View more about the median of this neighborhood's household incomes here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

In addition to the mean and the median, there is one other common measurement of the "typical" value of a data set: the **mode**.

Mode

The **mode** is the element of the data set that occurs most frequently.

The mode is fairly useless with data like weights or heights where there are a large number of possible values. The mode is most commonly used for categorical (or qualitative) data, for which median and mean cannot be computed.

Sometimes a data set may have more than one number that occurs most frequently. Then the data is considered to be *multi-modal*. Because a data set could have more than one mode, you do not want to use the Excel command for mode. It will only find the first mode in multi-modal data.

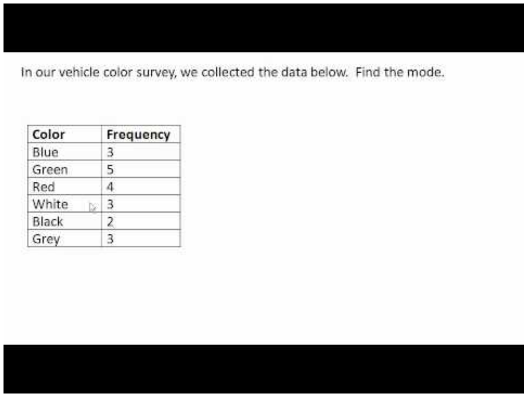
Example

In our vehicle color survey earlier in this section, we collected the data

Color	Frequency
Blue	3
Green	5
Red	4
White	3
Black	2
Grey	3

Which color is the mode?
[reveal-answer q="638793"]Show Solution[/reveal-answer]
[hidden-answer a="638793"]For this data, Green is the mode, since it is the data value that occurred the most frequently.[/hidden-answer]

Mode in this example is explained by the video here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

It is possible for a data set to have more than one mode if several categories have the same frequency, or no modes if each every category occurs only once.

Try It

Reviewers were asked to rate a product on a scale of 1 to 5. Find

1. The mean rating
2. The median rating
3. The mode rating

Rating	Frequency
1	4
2	8
3	7
4	3
5	1

Measures of Variation

Range and Standard Deviation

Consider these three sets of quiz scores:

Section A: 5 5 5 5 5 5 5 5 5 5

Section B: 0 0 0 0 0 10 10 10 10 10

Section C: 4 4 4 5 5 5 5 6 6 6

All three of these sets of data have a mean of 5 and median of 5, yet the sets of scores are clearly quite different. In section A, everyone had the same score; in section B half the class got no points and the other half got a perfect score, assuming this was a 10-point quiz. Section C was not as consistent as section A, but not as widely varied as section B.

In addition to the mean and median, which are measures of the "typical" or "middle" value, we also need a measure of how "spread out" or varied each data set is.

There are several ways to measure this "spread" of the data. The first is the simplest and is called the **range**.

Range

The range is the difference between the maximum value and the minimum value of the data set.

example

Using the quiz scores from above,

For section A, the range is 0 since both maximum and minimum are 5 and 5 – 5 = 0

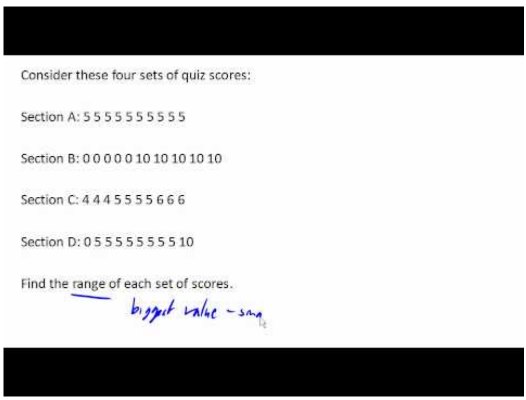
For section B, the range is 10 since 10 – 0 = 10

For section C, the range is 2 since 6 – 4 = 2

In the last example, the range seems to be revealing how spread out the data is. However, suppose we add a fourth section, Section D, with scores 0 5 5 5 5 5 5 5 5 10.

This section also has a mean and median of 5. The range is 10, yet this data set is quite different than Section B. To better illuminate the differences, we'll have to turn to more sophisticated measures of variation.

The range of this example is explained in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

Standard deviation

The standard deviation is a measure of variation based on measuring how far each data value deviates, or is different, from the mean. A few important characteristics:

- Standard deviation is always positive. Standard deviation will be zero if all the data values are equal, and will get larger as the data spreads out.
- Standard deviation has the same units as the original data.
- Standard deviation, like the mean, can be highly influenced by outliers.

Computing standard deviation by hand is a complex process, with many possibilities for little errors that can throw off your final answer. It is perfectly fine to use Excel to compute standard deviation. List all of your values in a column, like you would do to compute the mean or the median, but instead of those commands, use **=STDEV.S** or

=STDEV. Do not use =STDEV.P, which is meant for entire populations.

Here are the directions for computing standard deviation by hand. It is always good to try it once, or at least read through the process:

Using the data from section D, we could compute for each data value the difference between the data value and the mean:

data value	deviation: data value – mean
0	0-5 = -5
5	5-5 = 0
5	5-5 = 0
5	5-5 = 0
5	5-5 = 0
5	5-5 = 0
5	5-5 = 0
5	5-5 = 0
5	5-5 = 0
10	10-5 = 5

We would like to get an idea of the "average" deviation from the mean, but if we find the average of the values in the second column the negative and positive values cancel each other out (this will always happen), so to prevent this we square every value in the second column:

data value	deviation: data value – mean	deviation squared
0	0-5 = -5	$(-5)^2 = 25$
5	5-5 = 0	$0^2 = 0$
5	5-5 = 0	$0^2 = 0$
5	5-5 = 0	$0^2 = 0$
5	5-5 = 0	$0^2 = 0$
5	5-5 = 0	$0^2 = 0$
5	5-5 = 0	$0^2 = 0$
5	5-5 = 0	$0^2 = 0$
5	5-5 = 0	$0^2 = 0$
10	10-5 = 5	$(5)^2 = 25$

We then add the squared deviations up to get 25 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 25 = 50. Ordinarily we would then divide by the number of scores, n, (in this case, 10) to find the mean of the deviations. But we only do this if the data set represents a population; if the data set represents a sample (as it almost always does), we instead divide by n – 1 (in this case, 10 – 1 = 9).[1]

1. The reason we do this is highly technical, but we can see how it might be useful by considering the case of a small sample from a population that contains an outlier, which would increase the

So in our example, we would have 50/10 = 5 if section D represents a population and 50/9 = about 5.56 if section D represents a sample. These values (5 and 5.56) are called, respectively, the **population variance** and the **sample variance** for section D.

Variance can be a useful statistical concept, but note that the units of variance in this instance would be points-squared since we squared all of the deviations. What are points-squared? Good question. We would rather deal with the units we started with (points in this case), so to convert back we take the square root and get:

---

average deviation: the outlier very likely won't be included in the sample, so the mean deviation of the sample would underestimate the mean deviation of the population; thus we divide by a slightly smaller number to get a slightly bigger average deviation.

(1)

$$\text{population standard deviation} = \sqrt{\frac{50}{10}} = \sqrt{5} \approx 2.2$$

or

$$\text{sample standard deviation} = \sqrt{\frac{50}{9}} \approx 2.4$$

If we are unsure whether the data set is a sample or a population, we will usually assume it is a sample, and we will round answers to one more decimal place than the original data, as we have done above.

To compute standard deviation by hand

1. Find the deviation of each data from the mean. In other words, subtract the mean from the data value.

2. Square each deviation.
3. Add the squared deviations.
4. Divide by n, the number of data values, if the data represents a whole population; divide by $n - 1$ if the data is from a sample.
5. Compute the square root of the result.

example

Computing the standard deviation for Section B above, we first calculate that the mean is 5. Using a table can help keep track of your computations for the standard deviation:

data value	deviation: data value – mean	deviation squared
0	0-5 = -5	$(-5)^2 = 25$
0	0-5 = -5	$(-5)^2 = 25$
0	0-5 = -5	$(-5)^2 = 25$
0	0-5 = -5	$(-5)^2 = 25$
0	0-5 = -5	$(-5)^2 = 25$
10	10-5 = 5	$(5)^2 = 25$
10	10-5 = 5	$(5)^2 = 25$
10	10-5 = 5	$(5)^2 = 25$
10	10-5 = 5	$(5)^2 = 25$
10	10-5 = 5	$(5)^2 = 25$

Assuming this data represents a population, we will add the squared deviations, divide by 10, the number of data values, and compute the square root:

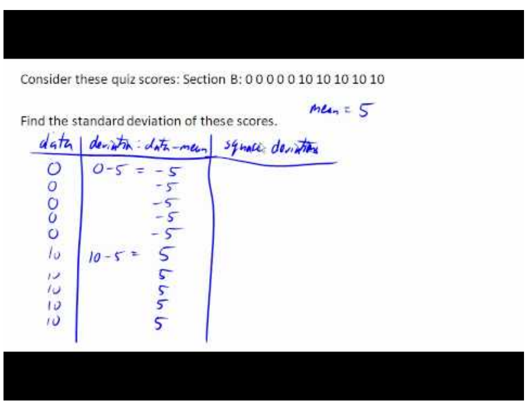
$$\sqrt{\frac{25+25+25+25+25+25+25+25+25+25}{10}} = \sqrt{\frac{250}{10}} = 5$$

Notice that the standard deviation of this data set is much larger than that of section D since the data in this set is more spread out.

For comparison, the standard deviations of all four sections are:

Section A: 5 5 5 5 5 5 5 5 5 5	Standard deviation: 0
Section B: 0 0 0 0 0 10 10 10 10 10	Standard deviation: 5
Section C: 4 4 4 5 5 5 5 6 6 6	Standard deviation: 0.8
Section D: 0 5 5 5 5 5 5 5 5 10	Standard deviation: 2.2

See the following video for more about calculating the standard deviation in this example by hand. Again, it is perfectly okay to use Excel to find the standard deviation.



A YouTube element has been excluded from this

version of the text. You can view it online here:
https://granite.pressbooks.pub/math502/?p=155

Try It

The price of a jar of peanut butter at 5 stores was $3.29, $3.59, $3.79, $3.75, and $3.99. Find the standard deviation of the prices.

Where standard deviation is a measure of variation based on the mean, **quartiles** are based on the median.

There are a number of different ways to compute quartiles, and each method results, sometimes, in slightly different answers. Excel does have a command for quartiles, but you will not always get the same answer as you will following the procedure below, so this is one instance where you will want to follow the instructions given here.

Quartiles

Quartiles are values that divide the data in quarters.

The first quartile (Q1) is the value so that 25% of the data values are below it; the third quartile (Q3) is the value so that 75% of the data values are below it. You may have guessed that the second quartile is the same as the median, since the median is the value so that 50% of the data values are below it.

This divides the data into quarters; 25% of the data is between the minimum and Q1, 25% is between Q1 and the median, 25% is between the median and Q3, and 25% is between Q3 and the maximum value.

While quartiles are not a 1-number summary of variation like standard deviation, the quartiles are used with the median, minimum, and maximum values to form a **5 number summary** of the data.

Five number summary

The five number summary takes this form:

Minimum, Q1, Median, Q3, Maximum

To find the first quartile, we need to find the data value so that 25% of the data is below it. If n is the number of data values, we compute a locator by finding 25% of n. If this locator is a decimal value, we round up, and find the data value in that position. If the locator is a whole number, we find the mean of the data value in that position and the next data value. This is identical to the process we used to find the median, except we use 25% of the data values rather than half the data values as the locator.

To find the first quartile, Q1

1. Begin by ordering the data from smallest to largest

2. Compute the locator: $L = 0.25n$
3. If L is a decimal value:
    - Round up to $L+$
    - Use the data value in the $L+$th position
4. If L is a whole number:
    - Find the mean of the data values in the Lth and $L+1$th positions.

To find the third quartile, Q3

Use the same procedure as for Q1, but with locator: $L = 0.75n$

Examples should help make this clearer.

examples

Suppose we have measured 9 females, and their heights (in inches) sorted from smallest to largest are:

59 60 62 64 66 67 69 70 72

What are the first and third quartiles?

Solution:

To find the first quartile we first compute the locator: 25% of 9 is L = 0.25(9) = 2.25. Since this value is not a whole number, we round up to 3. The first quartile will be the third data value: 62 inches.

To find the third quartile, we again compute the locator: 75% of 9 is 0.75(9) = 6.75. Since this value is not a whole number, we round up to 7. The third quartile will be the seventh data value: 69 inches.

Suppose we had measured 8 females, and their heights (in inches) sorted from smallest to largest are:

59 60 62 64 66 67 69 70

What are the first and third quartiles? What is the 5 number summary?

Solution:

To find the first quartile we first compute the locator: 25% of 8 is L = 0.25(8) = 2. Since this value *is* a whole number, we will find the mean of the 2nd and 3rd data values: (60+62)/2 = 61, so the first quartile is 61 inches.

The third quartile is computed similarly, using 75% instead of 25%. L = 0.75(8) = 6. This is a whole number, so we will find the mean of the 6th and 7th data values: (67+69)/2 = 68, so Q3 is 68.

Note that the median could be computed the same way, using 50%.

The 5-number summary combines the first and third quartile with the minimum, median, and maximum values.

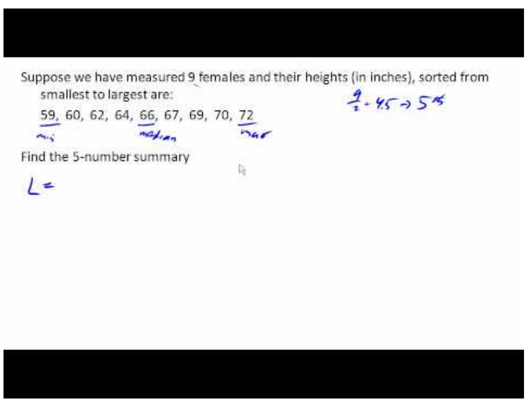
What are the 5-number summaries for each of the previous 2 examples?

Solution:

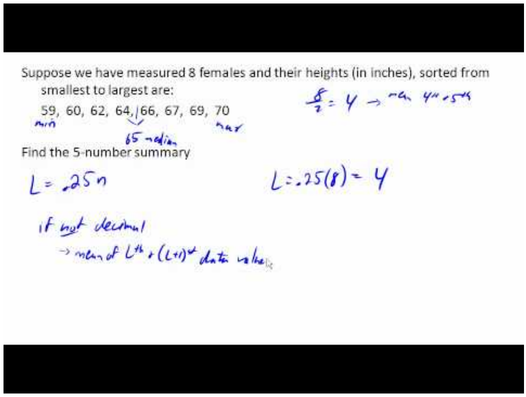
For the 9 female sample, the median is 66, the minimum is 59, and the maximum is 72. The 5 number summary is: 59, 62, 66, 69, 72.

For the 8 female sample, the median is 65, the minimum is 59, and the maximum is 70, so the 5 number summary would be: 59, 61, 65, 68, 70.

More about each set of women's heights is in the following videos.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155



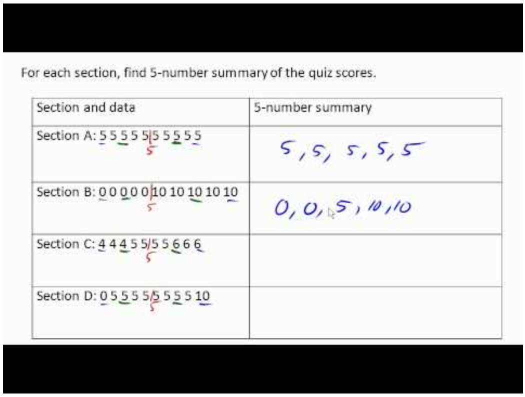
A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

Returning to our quiz score data: in each case, the first quartile locator is 0.25(10) = 2.5, so the first quartile will be the 3rd data value, and the third quartile will be the 8th data value. Creating the five-number summaries:

Section and data	5-number summary
Section A: 5 5 5 5 5 5 5 5 5 5	5, 5, 5, 5, 5
Section B: 0 0 0 0 0 10 10 10 10 10	0, 0, 5, 10, 10
Section C: 4 4 4 5 5 5 5 6 6 6	4, 4, 5, 6, 6
Section D: 0 5 5 5 5 5 5 5 5 10	0, 5, 5, 5, 10

Of course, with a relatively small data set, finding a five-number summary is a bit silly, since the summary contains almost as many values as the original data.

A video walkthrough of this example is available below.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

Try It

The total cost of textbooks for the term was collected from 36 students. Find the 5 number summary of this data.

$140 $160 $160 $165 $180 $220 $235 $240 $250 $260 $280 $285

$285 $285 $290 $300 $300 $305 $310 $310 $315 $315 $320 $320

$330 $340 $345 $350 $355 $360 $360 $380 $395 $420 $460 $460

Example

Returning to the household income data from earlier in the section, create the five-number summary.

Income (thousands of dollars)	Frequency
15	6
20	8
25	11
30	17
35	19
40	20
45	12
50	7

Solution:

By adding the frequencies, we can see there are 100 data values represented in the table. In Example 20, we found the median was $35 thousand. We can see in the table that the minimum income is $15 thousand, and the maximum is $50 thousand.

To find Q1, we calculate the locator: L = 0.25(100) = 25. This is a whole number, so Q1 will be the mean of the 25th and 26th data values.

Counting up in the data as we did before,

There are 6 data values of $15, so Values 1 to 6 are $15 thousand

The next 8 data values are \$20, so Values 7 to (6+8)=14 are \$20 thousand

The next 11 data values are \$25, so Values 15 to (14+11)=25 are \$25 thousand

The next 17 data values are \$30, so Values 26 to (25+17)=42 are \$30 thousand

The 25th data value is \$25 thousand, and the 26th data value is \$30 thousand, so Q1 will be the mean of these: (25 + 30)/2 = \$27.5 thousand.

To find Q3, we calculate the locator: L = 0.75(100) = 75. This is a whole number, so Q3 will be the mean of the 75th and 76th data values. Continuing our counting from earlier,

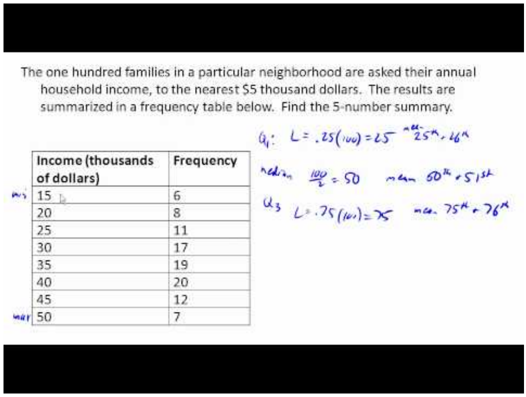
The next 19 data values are \$35, so Values 43 to (42+19)=61 are \$35 thousand

The next 20 data values are \$40, so Values 61 to (61+20)=81 are \$40 thousand

Both the 75th and 76th data values lie in this group, so Q3 will be \$40 thousand.

Putting these values together into a five-number summary, we get: 15, 27.5, 35, 40, 50

This example is demonstrated in this video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=155

Note that the 5 number summary divides the data into four intervals, each of which will contain about 25% of the data. In the previous example, that means about 25% of households have income between $40 thousand and $50 thousand.

For visualizing data, there is a graphical representation of a 5-number summary called a **box plot**, or box and whisker graph.

Box plot

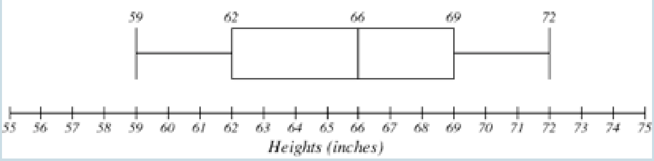
A **box plot** is a graphical representation of a five-number summary.

To create a box plot, a number line is first drawn. A box is drawn from the first quartile to the third quartile, and a line is drawn through the box at the median. "Whiskers" are extended out to the minimum and maximum values.

examples

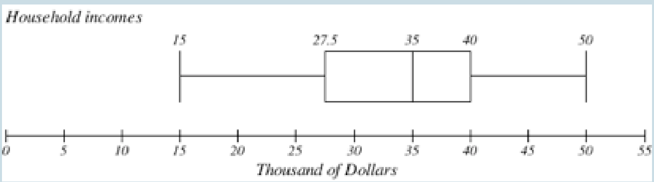
The box plot below is based on the 9 female height data with 5 number summary:

59, 62, 66, 69, 72.



The box plot below is based on the household income data with 5 number summary:

15, 27.5, 35, 40, 50



Box plot creation is described further here.

[Click here to view this video.]

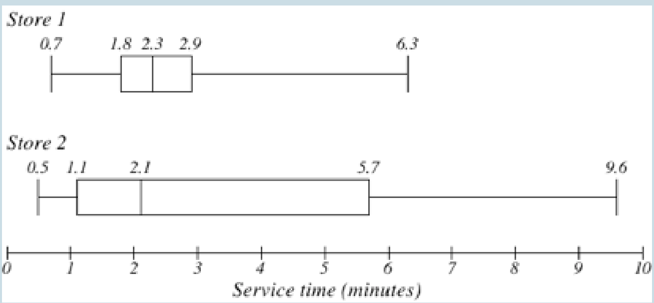
Try It

Create a box plot based on the textbook price data from the last Try It.

Box plots are particularly useful for comparing data from two populations.

examples

The box plot of service times for two fast-food restaurants is shown below.



While store 2 had a slightly shorter median service time (2.1 minutes vs. 2.3 minutes), store 2 is less consistent, with a wider spread of the data.

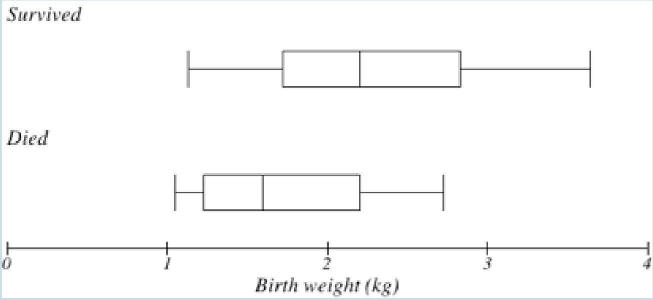
At store 1, 75% of customers were served within 2.9 minutes, while at store 2, 75% of customers were served within 5.7 minutes.

Which store should you go to in a hurry?

Solution:

That depends upon your opinions about luck – 25% of customers at store 2 had to wait between 5.7 and 9.6 minutes.

The box plot below is based on the birth weights of infants with severe idiopathic respiratory distress syndrome (SIRDS)[2]. The box plot is separated to show the birth weights of infants who survived and those that did not.



Comparing the two groups, the box plot reveals that the birth weights of the infants that died appear to be, overall, smaller than the weights of infants that survived. In fact, we can see that the median birth

---

2. van Vliet, P.K. and Gupta, J.M. (1973) Sodium bicarbonate in idiopathic respiratory distress syndrome. Arch. Disease in Childhood, 48, 249–255. As quoted on http://openlearn.open.ac.uk/mod/oucontent/view.php?id=398296&section=1.1.3

weight of infants that survived is the same as the third quartile of the infants that died.

Similarly, we can see that the first quartile of the survivors is larger than the median weight of those that died, meaning that over 75% of the survivors had a birth weight larger than the median birth weight of those that died.

Looking at the maximum value for those that died and the third quartile of the survivors, we can see that over 25% of the survivors had birth weights higher than the heaviest infant that died.

The box plot gives us a quick, albeit informal, way to determine that birth weight is quite likely linked to survival of infants with SIRDS.

The following video analyzes the examples above.

Click here to view this video.

Attributions

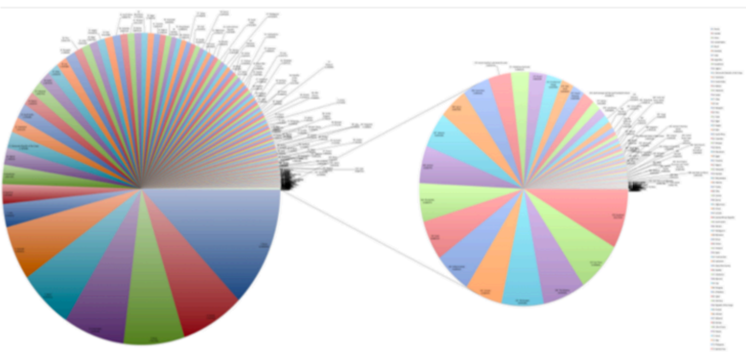
PUTTING IT TOGETHER: STATISTICS: DESCRIBING DATA

The following is excerpted from “The Trials and Tribulations of Data Visualization for Good” by Jake Porway.

The trials and tribulations of data visualization for good

“I love big data. It’s got such potential for storytelling.” At DataKind, we hear some version of this narrative every week. As more and more social organizations dip their toes into using data, invariably the conversation about data visualization comes up. There is a growing feeling that data visualization, with its combination of “engaging visuals” and “data-driven interactivity”, may be the magic bullet that turn opaque spreadsheets and dry statistics into funding, proof, and global action.

However, after four years of applying data-driven techniques to social challenges at DataKind, we feel that data visualization, while it does have an important place in our work, is a mere sliver of what it takes to work with data. Worse, the ubiquity of data visualization tools has lead to a wasteland of confusing, ugly, and sometimes unhelpful pie charts, word clouds, and worse.



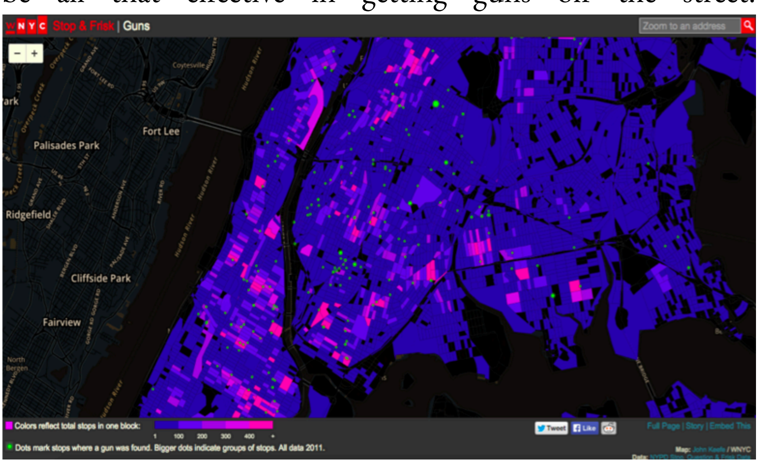
Ugh.

The challenge is that data visualization is not an end-goal, it is a process. It is often the final step in a long manufacturing chain along which data is poked, prodded, and molded to get to that pretty graph. Ignoring that process is at best misinformed, and at worst destructive.

Let me show you an example: In New York City, we had a very controversial program called Stop and Frisk that allowed police officers to stop people on the street they felt were a potential threat in an attempt to find and reclaim illegal weapons.

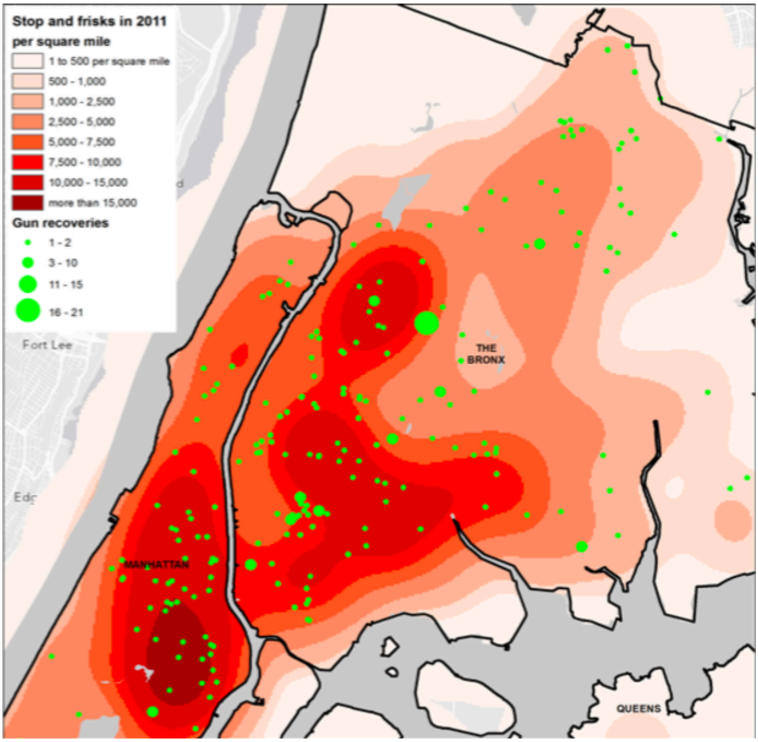
After a Freedom of Information Act (FOIA) request by the New York Civil Liberties Union (NYCLU) resulted in the New York Police Department (NYPD) releasing all of their Stop and Frisk data publicly, people flocked to the data to independently pick apart how effective the program was.

The figure below comes from WNYC, a public radio station located in New York City. Here they've shaded each city block brighter pink the more stops and frisks occurred there. The green dots on the map indicate where guns were found. What the figure shows is that the green dots do not appear as close to the hot pink squares as one would believe they should. The implication, then, is that Stop and Frisk may not actually be all that effective in getting guns off the street.



But then a citizen journalist created *this* map of the same data.

By simply changing the shading scheme slightly he notes that this map makes the green dots look much closer to the hot pink squares. In fact, he goes further to remove the artificial constraints of the block-by-block analysis and smooths over the whole area in New York, resulting in a map where those green dots stare unblinkingly on top of the hot-red stop and frisk regions.



The argument this author makes visually is that Stop and Frisk *does* in fact work.

So who's right here? Well both of them. And neither of them. These pictures are just that – pictures. Though they "use" data, they are not science. They are not analyses. They are mere visuals.

When data visualization is used simply to show alluring infographics about whether people like Coke or Pepsi better, the stakes of persuasion like this are low. But when they are used as arguments for or against public policy, the misuse of data visualization to persuade can have drastic consequences. **Data visualization without rigorous analysis is at best just rhetoric and, at worse, incredibly harmful.**

"Data for Humans vs. Data for Machines"

The fundamental challenge underlying this inadvertently malicious use of data comes, I believe, from a vagueness in terminology. When people crow about "the promise of data", they are often describing two totally different activities under the same umbrella. I've dubbed these two schools of thought "data for humans" vs. "data for machines".

Data for Humans: The most popular use of data, especially in the social sector, places all of the emphasis on the data itself as the savior. The idea is that, if we could just show people more data, we could prove our impact, encourage funding, and change behavior. Your bar charts, maps, and graphs pointing-up-and-to-the-right all fall squarely into this category. In fact almost all data visualization falls here, relying on the premise that showing a decisionmaker some data about the past will be all it takes to drive future change.

Unfortunately, while I believe data is a necessary part of this advocacy work, it is never sufficient by itself. The challenge with using "data for humans" is threefold:

1. **Humans don't make decisions based on data**, at least not alone**.** Plato once said "Human behavior flows from three main sources: desire, emotion, and knowledge." I want to believe he listed those aspects in that order

intentionally. Study after study has shown that humans rationalize beliefs with data, not vice versa. If behavior change were driven by data and graphs alone, we would be 50 years into a united battle against climate change. Conversely, we will leap to conclusions from data visualizations that "feel" right, but are not rigorously tested, like the conclusions from the Stop and Frisk images above.

2. **The public still treats data and data visualization as "fact" and "science".** I believe the public has gained enough visual literacy to question photojournalists or documentary filmmakers' motives, aware that theirs is an auteur behind the final piece that intends for us to walk away with their chosen understanding. We have yet to bring that same skepticism to data visualization, though we need to. The result of this illiteracy is that we are less critical of graphs and charts than written arguments because the use of data gives the sense that "fact" or "science" is at work, even if what we're doing is little more than visually bloviating.
3. **The data or visualization you see at the end of the road is opaque to interrogation.** It is difficult, if not impossible to know where that "58%" statistic or that flashy bar graph came from, grinning up at you from the page. Because we don't have ways to know how the data was collected, manipulated, and designed, we can't answer any of the questions we might want to raise

above. If point 2 means we need to treat data visualization as photojournalism, then this point implores us to go further to requiring forensic photographers in this work.

Data for Machines: For these reasons, DataKind specializes in projects focusing on what we refer to as “Data for Machines”. The promise of abundant data is not that we can show people more data, but that we can take advantage of computers, algorithms, and rigorous statistical methodologies to learn from these new datasets. The data is not the end goal, it is the raw resource we use to fuel computer systems that can learn from this information and, in many cases, even predict what is likely to happen in the future.

For example, instead of engaging in the Stop and Frisk gallery debate above, DataKind volunteers loaded the NYPD data into computers and created statistical models to rigorously test whether or not racial discrimination was occurring disproportionately in different parts of the city. While the models needed further evaluation, this analysis shows how data should be used. People shouldn’t try to draw conclusions from pictures of data – we’re notoriously bad at that as humans – we should be building models and using scientific methods to learn from data.

Celebrating Visualization

No surprise, creating data visualization well simply entails designing in a way that leads people to make scientific conclusions themselves.

There are many examples of insightful, persuasive, and downright clever data visualizations, but perhaps one of the best visualization practices I know of is to turn the idea of visualization on its head. Data visualization is incredibly good for allowing one to ask questions, not answer them. The huge amount of data that we have available to us now means that we need visual techniques just to help us make sense of what we need to try to make sense of.

So where do we go from here?

First off, you can boycott the tyranny of pie charts and word clouds, rail against those three pitfalls, and share these last two examples far and wide. But I think we can also all go out and start thinking about how data can truly be used to its fullest advantage. Aside from just using “data for machines,” the best data visualization should raise questions and inspire exploration, not just sum up information or try to tell us the answer. Today we have more information than ever before and we have a new opportunity to use it to mobilize others, provided we do so with sensitivity. Now, more than ever, we need to all be out there on the front lines looking beyond data visualization as merely a way to satisfy our funders' requirements and instead looking at data as a way to ask deep questions of our world and our future.

Attributions

EXERCISES

Exercises

Skills

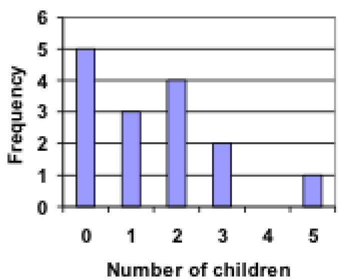
1. The table below shows scores on a Math test.
    a. Complete the frequency table for the Math test scores
    b. Construct a histogram of the data
    c. Construct a pie chart of the data

80	50	50	90	70	70	100	60	70	80	70	50
90	100	80	70	30	80	80	70	100	60	60	50

2. A group of adults where asked how many cars they had in their household
    a. Complete the frequency table for the car number data
    b. Construct a histogram of the data
    c. Construct a pie chart of the data

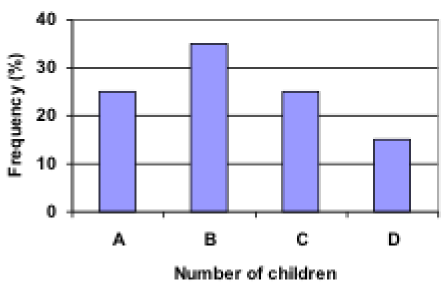
1	4	2	2	1	2	3	3	1	4	2	2
1	2	1	3	2	2	1	2	1	1	1	2

3. A group of adults were asked how many children they have in their families. The bar graph to the right shows the number of adults who indicated each number of children.
   a. How many adults where questioned?
   b. What percentage of the adults questioned had 0 children?

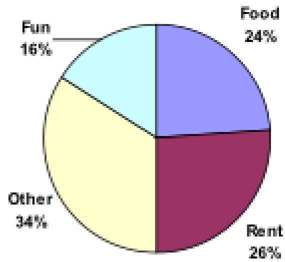


4. Jasmine was interested in how many days it would take an order from Netflix to arrive at her door. The graph below shows the data she collected.
   a. How many movies did she order?
   b. What percentage of the movies arrived in one day?

5. The bar graph below shows the *percentage* of students who received each letter grade on their last English paper. The class contains 20 students. What number of students earned an A on their paper?



6. Kori categorized her spending for this month into four categories: Rent, Food, Fun, and Other. The percents she spent in each category are pictured here. If she spent a total of $2600 this month, how much did she spend on rent?



7. A group of diners were asked how much they would pay for a meal. Their responses were: \$7.50, \$8.25, \$9.00, \$8.00, \$7.25, \$7.50, \$8.00, \$7.00.
    1. Find the mean
    2. Find the median
    3. Write the 5-number summary for this data
8. You recorded the time in seconds it took for 8 participants to solve a puzzle. The times were: 15.2, 18.8, 19.3, 19.7, 20.2, 21.8, 22.1, 29.4.
    1. Find the mean
    2. Find the median
    3. Write the 5-number summary for this data
9. Refer back to the histogram from question #3.
    1. Compute the mean number of children for the group surveyed
    2. Compute the median number of children for the group surveyed
    3. Write the 5-number summary for this data
    4. Create box plot
10. Refer back to the histogram from question #4.
    1. Computer the mean number of shipping days

2. Compute the median number of shipping days
3. Write the 5-number summary for this data
4. Create box plot

Concepts

11. The box plot below shows salaries for Actuaries and CPAs. Kendra makes the median salary for an Actuary. Kelsey makes the first quartile salary for a CPA. Who makes more money? How much more?

12. Referring to the boxplot above, what percentage of actuaries makes more than the median salary of a CPA?

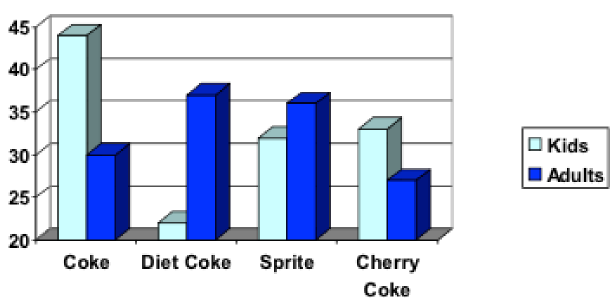
Exploration

13. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average length of time in months patients live once starting the treatment. Two researchers each follow a

different set of 40 AIDS patients from the start of treatment until their deaths. The following data (in months) are collected.Researcher 1: 3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34Researcher 2: 3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29

   a. Create comparative histograms of the data
   b. Create comparative boxplots of the data

14. A graph appears below showing the number of adults and children who prefer each type of soda. There were 130 adults and kids surveyed. Discuss some ways in which the graph below could be improved



15. Make up three data sets with 5 numbers each that have:
   a. the same mean but different standard deviations.
   b. the same mean but different medians.
   c. the same median but different means.
16. A sample of 30 distance scores measured in yards has a mean of 7, a variance of 16, and a standard deviation of

4.
    a. You want to convert all your distances from yards to feet, so you multiply each score in the sample by 3. What are the new mean, median, variance, and standard deviation?
    b. You then decide that you only want to look at the distance past a certain point. Thus, after multiplying the original scores by 3, you decide to subtract 4 feet from each of the scores. Now what are the new mean, median, variance, and standard deviation?

17. In your class, design a poll on a topic of interest to you and give it to the class.
    a. Summarize the data, computing the mean and five-number summary.
    b. Create a graphical representation of the data.
    c. Write several sentences about the topic, using your computed statistics as evidence in your writing.

Attributions

CHAPTER VI

CHAPTER 6: STATISTICS: NORMAL DISTRIBUTION

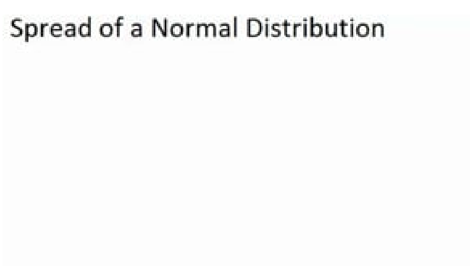
UNDERSTANDING NORMAL DISTRIBUTION

Objective

Here you will learn about the Normal Distribution. You will learn what it is and why it is important, and you will begin to develop an intuition for the rarity of a value in a set by comparing it to the mean and standard deviation of the data.

If you knew that the prices of t-shirts sold in an online shopping site were **normally distributed**, and had a mean cost of $10, with a standard deviation of $1.50, how could that information benefit you as you are looking at various t-shirt prices on the site? How could you use what you know if you were looking to make a profit by purchasing unusually inexpensive shirts to resell at prices that are more common?

Watch This: Spread of a Normal Distribution



A Vimeo element has been excluded from this

version of the text. You can view it online here:
https://granite.pressbooks.pub/math502/?p=319

Guidance

A **distribution** is an evaluation of the way that points in a data set are clustered or spread across their **range** of values. A **normal distribution** is a very specific symmetrical distribution that indicates, among other things, that exactly $\frac{1}{2}$ of the data is below the mean, and $\frac{1}{2}$ is above, that approximately 68% of the data is within 1, approximately 96% of the data is within 2, and approximately 99.7% is within 3 **standard deviations** of the mean.

There are a number of reasons that it is important to become familiar with the normal distribution, as you will discover throughout this chapter. Examples of values associated with normal distribution:

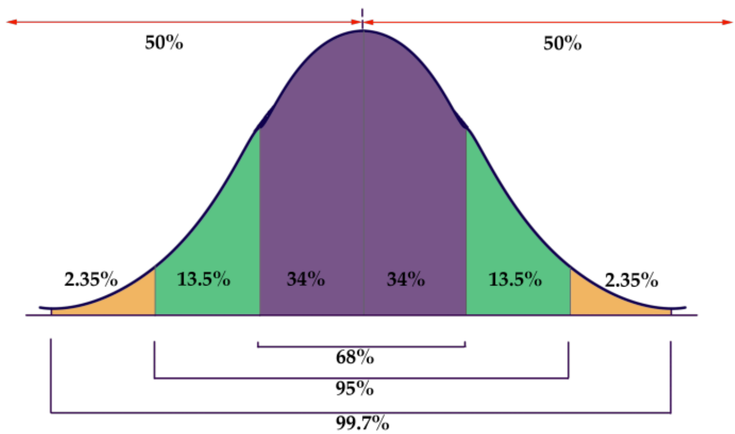
- Physical characteristics such as height, weight, arm or leg

length, etc.
- The percentile rankings of standardized testing such as the ACT and SAT
- The volume of water produced by a river on a monthly or yearly basis
- The velocity of molecules in an ideal gas

Knowing that the values in a set are exactly or approximately normally distributed allows you to get a feel for how common a particular value might be in that set. Because the values of a normal distribution are predictably clustered around the mean, you can estimate in short order the rarity of a given value in the set. In our upcoming lesson on the Empirical Rule, you will see that it is worth memorizing that normally distributed data has the characteristics mentioned above:

- 50% of all data points are above the mean and 50% are below
- Apx 68% of all data points are within 1 standard deviation of the mean
- Apx 95% of all data points are within 2 standard deviations of the mean
- Apx 99.7% of all data points are within 3 standard deviations of the mean

In this lesson, we will be practicing a 'rough estimate' of the probability that a value within a given range will occur in a particular set of data, just to develop an intuition of the use of a normal distribution. In subsequent lessons, we will become more specific with our estimates. The image below will be used in greater detail in the lesson on the Empirical Rule, but you may use it as a reference for this lesson also.



Example 1

Human height is commonly considered an approximately normally distributed measure. If the mean height of a male adult in the United States is 5' 1", with a standard deviation of 1.5", how common are men with heights greater than 6' 2"?

Solution

Since each standard deviation of this normally distributed data is 1.5", and 6' 2" is 400 above the mean for the population, 6' 2" is nearly 3 standard deviations above the mean. That tells us that men taller than 6' 2" are quite rare in this population.

Example 2

If the fuel mileage of a particular model of car is normally distributed, with a mean of 26 mpg and a standard deviation of 2 mpg, how common are cars with a fuel efficiency of 24 to 25 mpg?

Solution

We know that apx 68% of the cars in the population have an efficiency of between 24 and 28 mpg, since that would be 1 SD below and 1 SD above the mean. That suggests that apx 34% have an efficiency of 24 to 26 mpg, so we can say that it is uncommon to see a car with an efficiency between 24 and 25 mpg, but not extremely so.

Example 3

If the maximum jumping height of US high school high jumpers is normally distributed with a mean of 5' 11.5" and a SD of 2.2", how unusual is it to see a high school jumper clear 6' 3"?

Solution

If the mean is 5' 11.5", then 1 SD above is 6' 1.7" and 2 SDs is 6' 3.9". That means that less than 2.5% of jumpers 6' 3.9", so it would be pretty uncommon to see a high-school competitor exceed 6' 3".

Intro Problem Revisited

If you knew that the prices of t-shirts sold in an online shopping site were ***normally distributed****, and had a mean cost of $10, with a standard deviation of $1.50, how could that information benefit you as you are looking at various t-shirt styles and designs on the site? How could you use what you know if you were looking to make a profit by purchasing unusually inexpensive shirts to resell at prices that are more common?*

By knowing the mean and SD of the shirt prices, and knowing that they are normally distributed, you can estimate right away if a shirt is priced at a point significantly below the norm. For instance, with this data, we can estimate that a shirt priced at $7.00 is less expensive than apx 97.5% of all shirts on the site, and could likely be resold at a profit (assuming there is not something wrong the shirt that is not obvious from the listing).

Vocabulary

Distribution: an arrangement of values of a variable showing their observed or theoretical frequency of occurrence.

Range of values of a distribution: is the difference between the least and greatest values.

Normal distribution: a very specific distribution that is symmetric about its mean. Half the values of the random variable are below the mean and half are above the mean. Approximately 68% of the data is within 1 standard deviation of the mean; aproximately 96% is within 2 SDs, and 99.7% within 3 SDs.

Standard deviation: a measure of how spread out the data is from the mean. To determine if a data value is far from the mean, determine how many standard deviations it is from the mean. The SD is calculated as the square root of the variance.

Guided Practice

Assume the data to be normally distributed, and describe the rarity of an event using the following scale:

- 0% to < 1% probability = very rare
- 1% to < 5% = rare
- 5% to < 34% = uncommon
- 34% to < 50% = common
- 50% to 100% = likely

Questions

1. If the mean (μ) of the data is 75, and the standard deviation (σ) is 5, how common is a value between 70 and 75?
2. If the μ is .02 and the σ is .005, how common

is a value between .005 and .01?

3. If the μ is 1280 and the σ is 70, how common is a value between 1210 and 1350?
4. If the mean defect rate at a cellphone production plant is .1%, with a standard deviation of .03%, would it seem reasonable for a quality assurance manager to be concerned about 3 defective phones in a single 1000 unit run?

Solutions

1. A value of 70 is only 1 standard deviation below the mean, so a value between 70 and 75 would be expected approximately 34% of the time, so it would be common.
2. A value of .01 is 2 SDs below the mean, and .005 is 3 SDs below, so we would expect there to be about a 2.5% probability of a value occurring in that range. A value between 0.005 and 0.01 would be rare.
3. 1210 is 1 SD below the mean, and 1350 is 1 SD above the mean, so we would expect approximately 68% of the data to be in that range, meaning that it is likely that a value in that range would occur.

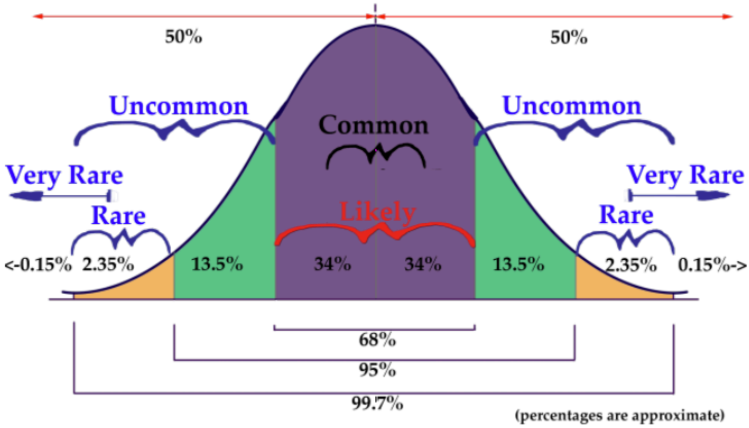
4. .1% translates into 1 per thousand, with a standard deviation of 3 per ten thousand. That means that 3 defects in the same thousand is nearly 7 SDs above the mean, well into the very rare category. While it is not impossible for random chance to result in such a value, it would certainly be prudent for the manager to investigate.

Practice Questions

Assume all sets/populations to be approximately normally distributed, and describe the rarity of an event using the following scale:

- 0% to< 1% probability = very rare
- 1% to < 5% = rare
- 5% to < 34% = uncommon
- 34% to < 50% = common
- 50% to 100% = likely.

You may reference the image below:



Attributions

This chapter contains material taken from *Math in Society* (on OpenTextBookStore) by David Lippman, and is used under a CC Attribution-Share Alike 3.0 United States (CC BY-SA 3.0 US) license.

This chapter contains material taken from of *Math for the Liberal Arts* (on Lumen Learning) by Lumen Learning, and is used under a *CC BY: Attribution* license.

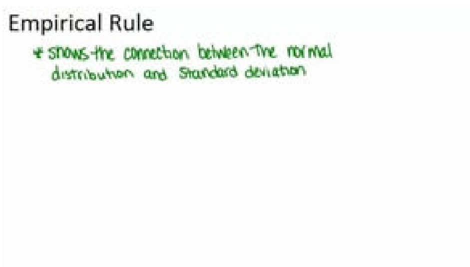
THE EMPIRICAL RULE

Objective

Here you will learn how to use the Empirical Rule to estimate the probability of an event.

If the price per pound of USDA Choice Beef is normally distributed with a mean of \$4.85/lb and a standard deviation of \$0.35/lb, what is the estimated probability that a randomly chosen sample (from a randomly chosen market) will be between \$5.20 and \$5.55 per pound?

Watch This: Empirical Rule

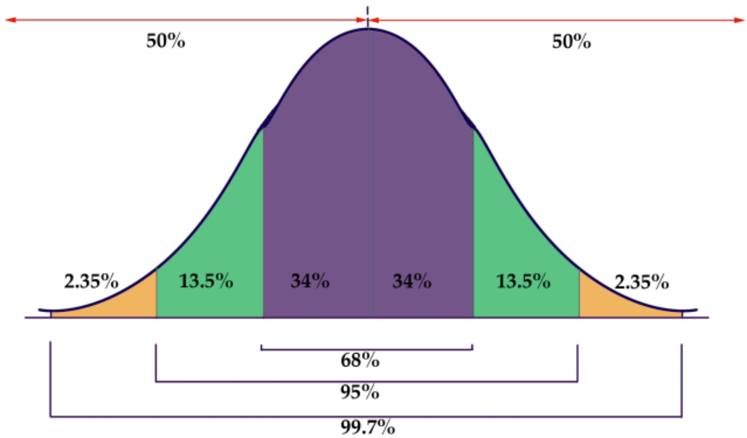


A Vimeo element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=321

Guidance

This reading on the **Empirical Rule** is an extension of the previous reading "Understanding the Normal Distribution." In the prior reading, the goal was to develop an intuition of the interaction between decreased probability and increased distance from the mean. In this reading, we will practice applying the Empirical Rule to estimate the specific probability of occurrence of a sample based on the range of the sample, measured in standard deviations.

The graphic below is a representation of the Empirical Rule:



The graphic is a rather concise summary of the **vital statistics** of a Normal Distribution. Note how the graph resembles a bell? Now you know why the normal distribution is also called a " bell curve."

- 50% of the data is above, and 50% below, the mean of the data
- Approximately 68% of the data occurs within 1 SD of the mean
- Approximately 95% occurs within 2 SD's of the mean
- Approximately 99.7% of the data occurs within 3 SDs of the mean

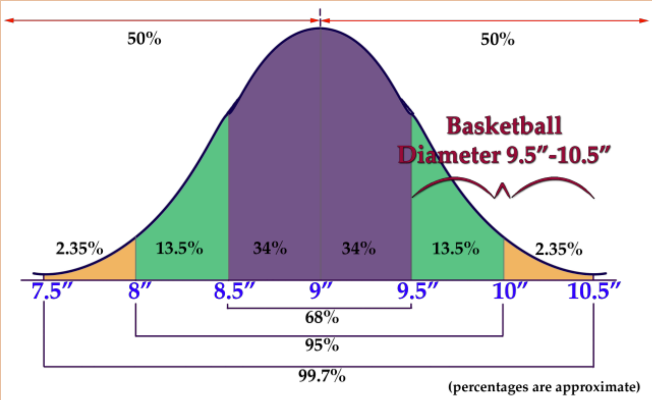
It is due to the probabilities associated with 1, 2, and 3 SDs that the Empirical Rule is also known as the **68–95–99.7 rule**.

Example 1

If the diameter of a basketball is normally distributed, with a mean (μ) of 9", and a standard deviation (σ) of 0.5", what is the probability that a randomly chosen basketball will have a diameter between 9.5" and 10.5"?

Solution

Since the σ = 0.5" and the μ = 9", we are evaluating the probability that a randomly chosen ball will have a diameter between 1 and 3 standard deviations above the mean. The graphic below shows the portion of the normal distribution included between 1 and 3 SDs:



The percentage of the data spanning the 2nd and 3rd SDs is 13.5% + 2.35% = 15.85%

The probability that a randomly chosen basketball will have a diameter between 9.5 and 10.5 inches is 15.85%.

Example 2

If the depth of the snow in my yard is normally distributed, with μ = 2.5" and **σ** = .25", what is the probability that a randomly chosen location will have a snow depth between 2.25 and 2.75 inches?

Solution

2.25 inches is $\mu - 1\sigma$, and 2.75 inches is $\mu + 1\sigma$, so the area encompassed approximately represents 34% + 34% = 68%.

The probability that a randomly chosen location will have a depth between 2.25 and 2.75 inches is 68%.

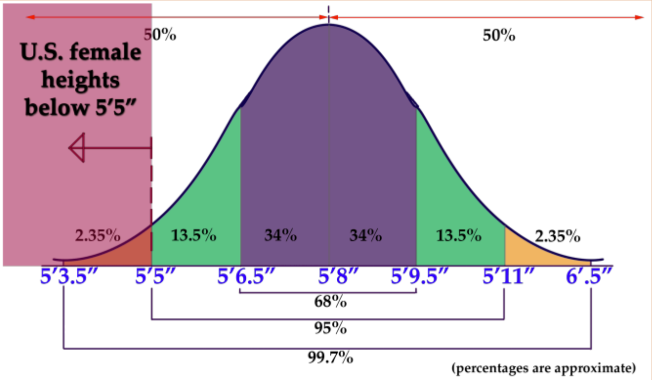
Example 3

If the height of women in the United States is normally distributed with μ = 5' 8" and σ = 1.5", what is the probability that a randomly chosen woman in the United States is shorter than 5' 5"?

Solution

This one is slightly different, since we aren't looking for the probability of a limited range of values. We want to evaluate the probability of a value occurring anywhere below 5' 5". Since the domain of a normal distribution is infinite, we can't actually state the probability of the portion of the distribution on "that end" because it has no "end"! What we need to do is add up the probabilities that we do know and subtract them from 100% to get the remainder.

Here is that normal distribution graphic again, with the height data inserted:



Recall that a normal distribution always has 50% of the data on each side of the mean. That indicates that 50% of US females are taller than 5' 8", and gives us a solid starting point to calculate from. There is another 34% between 5' 6.5" and 5' 8" and a final 13.5% between 5' 5" and 5' 6.5". Ultimately that totals: 50% + 34% + 13.5% = 97.5%. Since 97.5% of US females are 5' 5" or taller, that leaves 2.5% that are less than 5' 5" tall.

Intro Problem Revisited

If the price per pound of USDA Choice Beef is normally distributed with a mean of $4.85/lb and a standard deviation of $0.35/lb, what is the estimated probability that a randomly chosen sample (from a randomly chosen market) will be between $5.20 and $5.55 per pound?

$5.20 is $\mu + 1\sigma$, and $5.55 is $\mu + 2\sigma$, so the probability of a value occurring in that range is approximately 13.5%.

Vocabulary

Normal distribution: a common, but specific, distribution of data with a set of characteristics detailed in the lesson above.

Empirical Rule: a name for the way in which the normal distribution divides data by standard deviations: 68% within 1 SD, 95% within 2 SDs and 99.7 within 3 SDs of the mean

68-95-99.7 rule: another name for the Empirical Rule

Bell curve: the shape of a normal distribution

Guided Practice

1. A normally distributed data set has $\mu = 10$ and $\sigma = 2.5$, what is the probability of randomly selecting a value greater than 17.5 from the set?
2. A normally distributed data set has $\mu = .05$ and

σ = .01, what is the probability of randomly choosing a value between .05 and .07 from the set?

3. A normally distributed data set has μ = 514 and an unknown standard deviation, what is the probability that a randomly selected value will be less than 514?

Solutions

1. If μ = 10 and σ = 2.5, then 17.5 = μ + 3σ. Since we are looking for all data above that point, we need to subtract the probability that a value will occur below that value from 100%: The probability that a value will be less than 10 is 50%, since 10 is the mean. There is another 34% between 10 and 12.5, another 13.5% between 12.5 and 15, and a final 2.35% between 15 and 17.5. 100% −50% −34% −13.5% −2.35% = 0.15% probability of a value greater than 17.5
2. 0.05 is the mean, and 0.07 is 2 standard deviations above the mean, so the probability of a value in that range is 34% + 13.5% = 47.5%
3. 514 is the mean, so the probability of a value less than that is 50%.

Attributions

Z-SCORES

Objectives

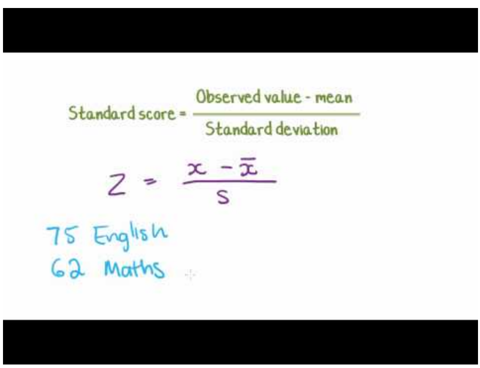
1. Here you will learn how z-scores can be used to evaluate how extreme a given value is in a particular set or population.
2. Here you will learn to evaluate z-scores as they relate to probability.
3. Here you will learn to calculate the probability of a z-score between two others.

Part I

Using the Empirical Rule can give you a good idea of the probability of occurrence of a value that happens to be exactly one, two or three to either side of the mean, but how do you compare the probabilities of values that are in between standard deviations?

Watch This: Maths Tutorial: Z scores

The British video below is very clear and easy to follow. It is worth noting, particularly for US students, that the instructor uses the notation x rather than μ for mean, and pronounces z as "zed."



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=323

Z-scores are related to the Empirical Rule from the standpoint of being a method of evaluating how extreme a particular value is in a given set. **You can think of a z-score as the number of standard deviations there are between a given value and the mean of the set.** While the Empirical Rule allows you to associate the first three standard deviations with the percentage of data that each SD includes, the z-score allows you to state (as accurately as you like), just how many SDs a given value is above or below the mean.

Conceptually, the z-score calculation is just what you might expect, given that you are calculating the number of SDs between a value and the mean. You calculate the z-score by first calculating the difference between your value and the mean, and then dividing that amount by the standard deviation of the set. The formula looks like this:

$$\text{z-score} = \frac{(\text{value} - \text{mean})}{\text{standard deviation}} = \frac{(x = \mu)}{\sigma}$$

In this lesson, we will practice calculating the z-score for various values. In the next lesson, we will learn how to associate the z-score of a value with the probability that the value will occur.

Example 1

What is the z-score of a value of 27, given a set mean of 24, and a standard deviation of 2?

Solution

To find the z-score we need to divide the difference between the value, 27, and the mean, 24, by the standard deviation of the set, 2.

$$\text{z-score} = \frac{27-\mu}{\sigma} = \frac{27-24}{2} = \frac{3}{2}$$

$$\text{z-score of } 27 = +1.5$$

This indicates that 27 is 1.5 standard deviations above the mean.

Example 2

What is the z-score of a value of 104.5, in a set with μ = 125 and σ = 6.2?

Solution

Find the difference between the given value and the mean, then divide it by the standard deviation.

$$\text{z-score} = \frac{104.5 - \mu}{\sigma} = \frac{104.5 - 125}{6.2} = \frac{-20.5}{6.2}$$

$$\text{z-score of } 104005 = -3.306$$

Note that the z-score is negative, since the measured value, 104.5, is less than (below) the mean, 125.

Example 3

Find the value represented by a z-score of 2.403, given μ = 63 and **σ** = 4.25.

Solution

This one requires that we solve for a missing value rather than for a missing z-score, so we just need to fill in our formula with what we know and solve for the missing value:

$$\text{z-score} = \frac{x-\mu}{\sigma}$$

$$2.403 = \frac{x-63}{4.25}$$

$$10.213 = x - 63$$

$$73.213 = x$$

73.213 has a z-score of 2.403

Intro Problem Revisited

Using the Empirical Rule can give you a good idea of the probability of occurrence of a value that happens to be right on one of the first three standard deviations to either side of the mean, but how do you compare the probabilities of values that are in between standard deviations?

The z-score of a value is the count of the number of standard deviations between the value and the mean of the set. You can find it by subtracting the value from the mean, and dividing the result by the standard deviation.

Vocabulary

The z-score of a value is the number of standard deviations between the value and the mean of the set.

Guided Practice

1. What is the z-score of the price of a pair of skis that cost \$247, if the mean ski price is \$279, with a standard deviation of \$16?
2. What is the z-score of a 5-scoop ice cream cone if the mean number of scoops is 3, with a standard deviation of 1 scoop?
3. What is the z-score of the weight of a cow that tips the scales at 825 lbs, if the mean

weight for cows of her type is 1150 lbs, with a standard deviation of 77 lbs?

4. What is the z-score of a measured value of 0.0034, given μ = 0.0041 and σ = 0.0008?

Solutions

1. First find the difference between the measured value and the mean, then divide that difference by the standard deviation:
$\frac{247-279}{16}=\frac{-32}{16}$

   z-score = −2
2. This one is easy: The difference between 5 scoops and 3 scoops is +2, and we divide that by the standard deviation of 1, so the z-score is +2.
3. First find the difference between the measured value and the mean, then divide that difference by the standard deviation:
$\frac{825\text{ lbs}-1150\text{ lbs}}{771\text{ lbs}}=\frac{-325}{77}$

   z-score = −4.2407
4. First find the difference between the measured value and the mean, then divide that difference by the standard deviation:

$$\frac{0.0034 - 0.0041}{0.0008} = \frac{-0.0007}{0.0008}$$

z-score = −0.875

Part II

Knowing the z-score of a given value is great, but what can you do with it? How does a z-score relate to probability? What is the probability of occurrence of a z-score less than +2.47?

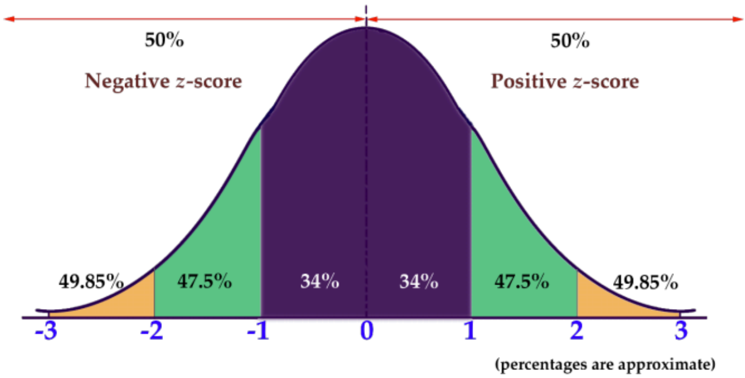
Watch This: The Normal Distribution

The video below provides a demonstration of how to use a z-score probability reference table, as we do in this lesson. The table he uses in the video is slightly different, but the concept is the same.

A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=323

Since z-scores are a measure of the number of SDs between a value and the mean, they can be used to calculate probability by comparing the location of the z-score to the area under a normal curve either to the left or right. The area can be calculated using calculus, but we will just use a table to look up the area.

I believe that the concept of comparing z-scores to probability is most easily understood with a graphic like the one we used in the lesson on the Empirical Rule, so I included one below. Be sure to review the examples to see how the scores work.



Like the graphic we viewed in the Empirical Rule lesson, this one only provides probability percentages for integer values of z-scores (standard deviations). In order to find the values for z-scores that aren't integers, you can use a table like the one below. To find the value associated with a given z-score, you find the first decimal of your z-score on the left or right side and then the 2nd decimal of your z-score across the top or bottom of the table. Where they intersect you will find the decimal expression of the percentage of values that are less than your sample (see example 4).

Table 1 (scroll to see all values)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.

Table 1 (scroll to see all values)

2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.
3.9	1	1	1	1	1	1	1	1	1
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.

Z-score tables like the one above describe the probability that a given value, or any value less than it, will occur in a given set. This particular table assumes you are looking to find the probability associated with a positive z-score. You may have additional work to do if the z-score is negative.

- To find the percentage of values greater than a negative Z score, just look up the matching positive Z score value.
- To find the percentage of values less than a negative z-score, subtract the chart value from 1.
- To find the percentage of values greater than a positive z-score, subtract the chart value from 1.

Example 4

What is the probability that a value with a z-score less than 2.47 will occur in a normal distribution?

Solution

Scroll up to the table above and find “2.4” on the left or right side. Now move across the table to “0.07” on the top or bottom, and record the value in the cell: **0.9932**. That tells us that **99.32% of values in the set are at or below a z-score of 2.47**.

Example 5

What is the probability that a value with a z-score greater than 1.53 will occur in a normal distribution?

Solution

Scroll up to the table of z-score probabilities again and find the intersection between 1.5 on the left or right and 3 on the top or bottom, record the value in the cell: **0.937**.

That decimal lets us know that 93.7% of values in the set are *below* the z-score of 1.53. To find the percentage that is above that value, we subtract 0.937 from 1.0 (or 93.7% from 100%), to get **0.063** or **6.3%**.

Example 6

What is the probability of a random selection being less than 3.65, given a normal distribution with $\mu = 5$ and $\sigma = 2.2$?

Solution

This question requires us to first find the z-score for the value 3.65, then calculate the percentage of values below
that z-score from a reference.

1. Find the z-score for 3.65, using the z-score formula:
$\frac{(x-\mu)}{\sigma} = \frac{3.65-5}{2.2} = \frac{-1.35}{2.2} \approx -0.61$
2. Now we can scroll up to our z-score reference above and find the intersection of 0.6 and 0.01, which should be .7291.
3. Since this is a negative z-score, and we want the percentage of values below it, we subtract that decimal from 1.0 (reference the three steps highlighted by bullet points below the chart if you didn't recall this), to get 1 – .7291 = .2709

There is approximately a 27.09% probability that a value less than 3.65 would occur from a random selection
of a normal distribution with mean 5 and standard deviation 2. 2.

Concept Problem Revisited

Knowing the z-score of a given value is great, but what can you do with it? How does a z-score relate to probability? What is the probability of occurrence of a z-score less than 2.47?

A z-score lets you calculate the probability that a randomly selected value will be greater or less than a particular value in a set.

To find the probability of a z-score below +2.47, using a reference such as the table in the lesson above:

1. Find 2.4 on the left or right side
2. Move across to 0.07 on the top or bottom.
3. The cell you arrive at says: 0.9932, which means that apx 99.32% of the values in a normal distribution will occur below a z-score of 2.47.

Vocabulary

Z-score table: a table that associates the various common z-scores between 0 and 3.99 with the decimal probability of being less than or equal to that z-score.

Guided Practice

1. What is the probability of occurrence of a value with z-score greater than 1.24?
2. What is the probability of $z < -.23$?
3. What is $P(Z < 2.13)$?

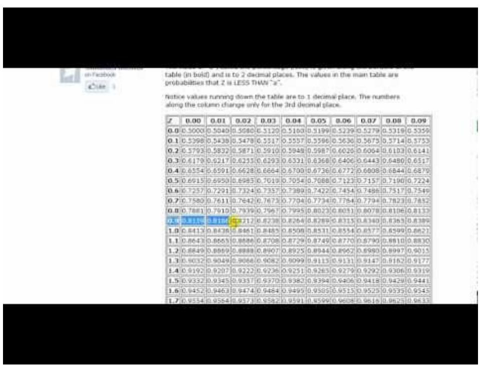
Solutions

1. Since this is a positive z-score, we can use the value for $z = 1.24$ directly from the table, and just express it as a percentage: 0.8925 or 89.25%
2. This is a negative z-score, and we want the percentage of values greater than it, so we need to subtract the value for $z = +0.23$ from 1: $1 - 0.591 = .409$ or 40.9%
3. This is a positive z-score, and we need the percentage of values below it, so we can use the percentage associated with $z = +2.13$ directly from the table: 0.9834 or 98.34%

Part III

Do z-score probabilities always need to be calculated as the chance of a value either above or below a given score? How would you calculate the probability of a z-score between −0.08 and +1.92?

Watch This: Reading Probabilities from the Z-table



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=323

To calculate the probability of getting a value with a z-score between two other z-scores, you can either use a reference table to look up the value for both scores and subtract them to find the difference, or you can use technology. In this lesson, which is an extension of Z-scores and Z-scores II, we will practice both methods.

Historically, it has been very common to use a z-score probability table like the one below to look up the probability associated with a given z-score:

Table 2 (scroll to see all values)

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.

Table 2 (scroll to see all values)

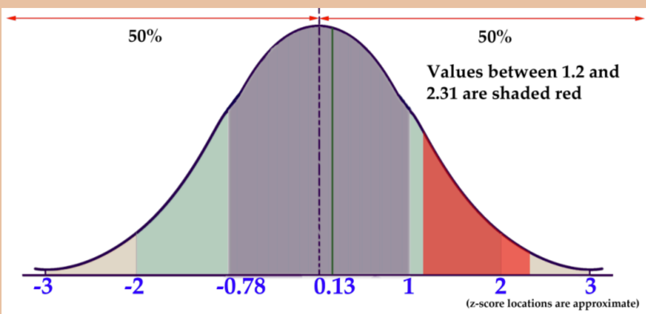
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.
3.9	1	1	1	1	1	1	1	1	1
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.

Example 7

What is the probability associated with a z-score between 1.2 and 2.31?

Solution

To evaluate the probability of a value occurring within a given range, you need to find the probability of both the upper and lower values in the range, and subtract to find the difference.



- First find z = 1.2 on the z-score probability reference above: .8849. Remember that value represents the percentage of values below 1.2.
- Next, find and record the value associated with z = 2.31: .9896
- Since approximately 88.49% of all values are

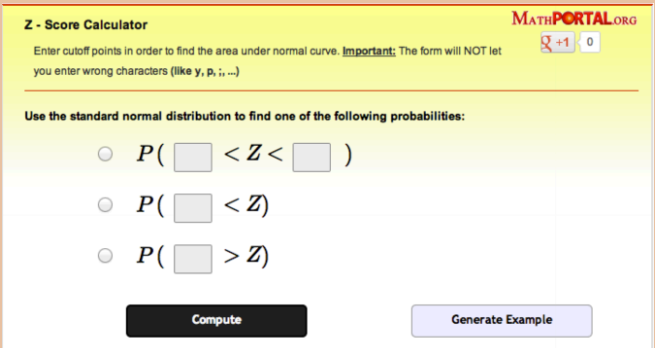
below z = 1.2 and approximately 98.96% of all values are below z = 2.31, there are 98.96% −88.49% = 10.47% of values between.

Example 8

What is the probability that a value with a z-score between -1.32 and +1.49 will occur in a normal distribution?

Solution

Let's use the online calculator at mathportal.org for this one. When you open the page, you should see a window like this:



All you need to do is select the radio button to the left of the first type of probability, input "−1.32" into the first box, and 1.49 into the second. When you click **Compute**, you should get the result $P(-1.32 < Z < 1.49) = 0.8385$

Which tells us that there is approximately and 83.85% probability that a value with a z-score between 1.32 and 1.49 will occur in a normal distribution.

Notice that the calculator also details the steps involved with finding the answer:

1. Estimate the probability using a graph, so you have an idea of what your answer should be.
2. Find the probability of $z < 1.49$, using a reference. (0.9319)

3. Find the probability of $z < -1.32$, again, using a reference. (0.0934)
4. Subtract the values: 0.9319 – 0.0934 = 0.8385 or 83.85%

Example 9

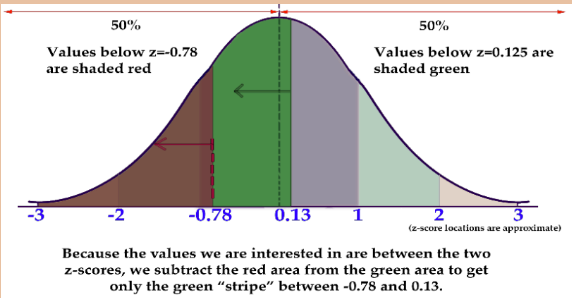
What is the probability that a random selection will be between 8.45 and 10.25, if it is from a normal distribution with $\mu = 10$ and $\sigma = 2$?

Solution

This question requires us to first find the z-scores for the value 8.45 and 10.25, then calculate the percentage of value between them by using values from a z-score reference and finding the difference.

1. Find the z-score for 8.45, using the z-score formula: $\frac{(x-\mu)}{\sigma}$
$\frac{8.45-10}{2} = \frac{-1.55}{2} \approx -0.78$

2. Find the z-score for 10.25 the same way:
$\frac{10.25-10}{2} = \frac{0.25}{2} \approx .13$

3. Now find the percentages for each, using a reference (don't forget we want the probability of values less than our negative score and less than our positive score, so we can find the values between):
$P(Z < -0.78) = .2177$ or 21.77%
$P(Z < .13) = .5517$ or 55.17%
4. At this point, let's sketch the graph to get an idea what we are looking for:


5. Finally, subtract the values to find the difference:
$.5517 - .2177 = .3340$ *or about* 33.4%
There is approximately a 33.4% probability that a value between 8.45 and 10.25 would result from a random selection of a normal distribution with mean 10 and standard

deviation 2.

Concept Problem Revisited

Do z-score probabilities always need to be calculated as the chance of a value either above or below a given score? How would you calculate the probability of a z-score between −0.08 and +1.92?

After this lesson, you should know without question that z-score probabilities do not need to assume only probabilities above or below a given value, the probability between values can also be calculated.

The probability of a z-score below −0.08 is 46.81%, and the probability of a z-score below 1.92 is 97.26%, so the probability between them is 97.26%−46.81% = 50.45%.

Vocabulary

z-score: a measure of how many standard deviations there are between a data value and the mean.

z-score probability table: a table that associates z-scores to area under the normal curve. The table may be used to associate a Z-score with a percent probability.

Guided Practice

1. What is the probability of a z-score between −0.93 and 2.11?
2. What is $P(1.39 < Z < 2.03)$?
3. What is $P(-2.11 < Z < 2.11)$?

Solutions

1. Using the z-score probability table above, we can see that the probability of a value below −0.93 is .1762, and the probability of a value below 2.11 is .9826. Therefore, the probability of a value between them is .9826−.1762 = .8064 or 80.64%
2. Using the z-score probability table, we see that the probability of a value below z = 1.39 is

.9177, and a value below z = 2.03 is .9788. That means that the probability of a value between them is .9788 – .9177 = .0611 or 6.11%

3. The probability of a z-score between –2.11 and +2.11 is about 96.52%.

Attributions

This chapter contains material taken from *Math in Society* (on OpenTextBookStore) by David Lippman, and is used under a CC Attribution-Share Alike 3.0 United States (CC BY-SA 3.0 US) license.

This chapter contains material taken from of *Math for the Liberal Arts* (on Lumen Learning) by Lumen Learning, and is used under a *CC BY: Attribution* license.

EXERCISES

Review Exercises

1. Scores on a certain standardized test have a mean of 500, and a standard deviation of 100. How common is a score between 600 and 700?
2. Considering a full-grown show-quality male Siberian Husky has a mean weight of 52.5 lbs, with SD of 7.5 lbs, how common are male huskies in the 37.5–45 lbs range?
3. A population $\mu = 125$, and $\sigma = 25$, how common are values in the 100 – 150 range?
4. Population $\mu = 0.0025$ and $\sigma = 0.0005$, how common are values between 0.0025 and 0.0030?
5. A 12 oz can of soda has a mean volume of 12 oz, with a standard deviation of .25 oz. How common are cans with between 11 and 11.5 oz of soda?
6. $\mu = 0.0025$ and $\sigma = 0.0005$, how common are values between 0.0045 and 0.005?
7. If a population $\mu = 1130$ and $\sigma = 5$, how common are values between 0 and 1100?
8. Assuming population $\mu = 1130$ and $\sigma = 5$, how common are values between 1125 and 1135?

9. The American Robin Redbreast has a mean weight of 77 g, with a standard deviation of 6 g. How common are Robins in the 59 g–71 g range?
10. Population $\mu = 0.25\%$ and $\sigma = 0.05\%$, how common are values between 0.35% and 0.45%?
11. Population $\mu = 156.5$ and $\sigma = 0.25$, how common are values between 155 and 156?

Assume all distributions to be normal or approximately normal, and calculate percentages using the 68–95–99.7 rule.

12. Given mean 63 and standard deviation of 168, find the approximate percentage of the distribution that lies between −105 and 567.
13. Approximately what percent of a normal distribution is between 2 standard deviations and 3 standard deviations from the mean?
14. Given standard deviation of 74 and mean of 124, approximately what percentage of the values are greater than 198?
15. Given $\sigma = 39$ and $\mu = 101$, approximately what percentage of the values are less than 23?
16. Given mean 92 and standard deviation 189, find the approximate percentage of the distribution that lies between −286 and 470.
17. Approximately what percent of a normal

distribution lies between $\mu + 1\sigma$ and $\mu + 2\sigma$?

18. Given standard deviation of 113 and mean 81, approximately what percentage of the values are less than −145?
19. Given mean 23 and standard deviation 157, find the approximate percentage of the distribution that lies between 23 and 337.
20. Given $\sigma = 3$ and $\mu = 84$, approximately what percentage of the values are greater than 90?
21. Approximately what percent of a normal distribution is between μ and $\mu+1\sigma$?
22. Given mean 118 and standard deviation 145, find the approximate percentage of the distribution that lies between −27 and 118.
23. Given standard deviation of 81 and mean 67, approximately what percentage of values are greater than 310?
24. Approximately what percent of a normal distribution is less than 2 standard deviations from the mean?
25. Given $\mu + 1\sigma = 247$ and $\mu + 2\sigma = 428$, find the approximate percentage of the distribution that lies between 66 and 428.
26. Given $\mu - 1\sigma = -131$ and $\mu + 1\sigma = 233$, approximately what percentage of the values are greater than −495?

The following problems are based on section 6.3 in your text.

27. Given a distribution with a mean of 70 and standard deviation of 62, find a value with a z-score of -1.82.
28. What does a z-score of 3.4 mean?
29. Given a distribution with a mean of 60 and standard deviation of 98, find the z-score of 120.76.
30. Given a distribution with a mean of 60 and standard deviation of 21, find a value with a z-score of 2.19.
31. Find the z-score of 187.37, given a distribution with a mean of 185 and standard deviation of 1.
32. What does a z-score of -3.8 mean?
33. Find the z-score of 125.18, given a distribution with a mean of 101 and standard deviation of 62.
34. Given a distribution with a mean of 117 and standard deviation of 42, find a value with a z-score of -0.94.
35. Given a distribution with a mean of 126 and standard deviation of 100, find a value with a z-score of -0.75.
36. Find the z-score of 264.16, given $\mu = 188$ and

$\sigma = 64$.

37. Find a value with a z-score of -0.2, given $\mu = 145$ and $\sigma = 56$.
38. Find the z-score of 89.79 given $\mu = 10$ and $\sigma = 79$.
39. Find the probabilities, use the table from the lesson above.
40. What is the probability of a z-score less than +2.02?
41. What is the probability of a z-score greater than +2.02?
42. What is the probability of a z-score less than −1.97?
43. What is the probability of a z-score greater than −1.97?
44. What is the probability of a z-score less than +0.09?
45. What is the probability of a z-score less than −0.02?
46. What is $P(Z < 1.71)$?
47. What is $P(Z > 2.22)$?
48. What is $P(Z < -1.19)$?
49. What is $P(Z > -2.71)$?
50. What is $P(Z < 3.71)$?
51. What is the probability of the random occurrence of a value greater than 56 from a normally distributed population with mean

62 and standard deviation 4.5?

52. What is the probability of a value of 329 or greater, assuming a normally distributed set with mean 290 and standard deviation 32?
53. What is the probability of getting a value below 1.2 from the random output of a normally distributed set with $\mu = 2.6$ and $\sigma = .9$?
54. Find the probabilities, use the table from the lesson or an online resource.
55. What is the probability of a z-score between +1.99 and +2.02?
56. What is the probability of a z-score between −1.99 and +2.02?
57. What is the probability of a z-score between −1.20 and −1.97?
58. What is the probability of a z-score between +2.33 and −0.97?
59. What is the probability of a z-score greater than +0.09?
60. What is the probability of a z-score greater than −0.02?
61. What is $P(1.42 < Z < 2.01)$?
62. What is $P(1.77 < Z < 2.22)$?
63. What is $P(-2.33 < Z < -1.19)$?
64. What is $P(-3.01 < Z < -0.71)$?
65. What is $P(2.66 < Z < 3.71)$?

66. What is the probability of the random occurrence of a value between 56 and 61 from a normally distributed population with mean 62 and standard deviation 4.5?
67. What is the probability of a value between 301 and 329, assuming a normally distributed set with mean 290 and standard deviation 32?

Attributions

CHAPTER VII

CHAPTER 7: PROBABILITY

WHY IT MATTERS: PROBABILITY

Why learn how to compute probability?

According to the news, the lottery jackpot is climbing by the hour. Long lines of dreamers are forming wherever lottery tickets are sold. While you don't usually buy lottery tickets, it is getting tempting. Just imagine what you could do with $100 million dollars. Perhaps you could retire early or never even go to work. Maybe buy a rare fancy car. All you need to do is pick the correct numbers, and the jackpot is all yours!

It sounds easy enough; just six simple numbers. But how likely are you to win? And could you increase the likelihood of winning by purchasing more lottery tickets?

To answer these questions, you need to know about permutations and combinations. So learn about them as you complete this module, and then we'll return to the lottery at the end. Then you'll be able to decide whether you want to stand in line to purchase a ticket.

Attributions

COMPUTING THE PROBABILITY OF AN EVENT

Learning Outcomes

- Describe a sample space and simple and compound events in it using standard notation
- Calculate the probability of an event using standard notation
- Calculate the probability of two independent events using standard notation
- Recognize when two events are mutually exclusive
- Calculate a conditional probability using standard notation

Probability is the likelihood of a particular outcome or event happening. Statisticians and actuaries use probability to make predictions about events. An actuary that works for a car insurance company would, for example, be interested in how likely a 17 year old male would be to get in a car accident. They would use data from past events to make predictions about future events using the characteristics of probabilities, then use this information to calculate an insurance rate.

In this section, we will explore the definition of an event, and learn how to calculate the probability of it's occurance. We will also practice using standard mathematical notation to calculate and describe different kinds of probabilities.

Basic Concepts

If you roll a die, pick a card from deck of playing cards, or randomly select a person and observe their hair color, we are executing an experiment or procedure. In probability, we look at the likelihood of different outcomes.

We begin with some terminology.

Events and Outcomes

- The result of an experiment is called an **outcome**.
- An **event** is any particular outcome or group of outcomes.
- A **simple event** is an event that cannot be broken down further

- The **sample space** is the set of all possible simple events.

example

If we roll a standard 6-sided die, describe the sample space and some simple events.

Solution:

The sample space is the set of all possible simple events: {1,2,3,4,5,6}

Some examples of simple events:

- We roll a 1
- We roll a 5

Some compound events:

- We roll a number bigger than 4
- We roll an even number

Basic Probability

Given that all outcomes are equally likely, we can compute the probability of an event E using this formula:

$$P(E) = \frac{\text{Number of outcomes corresponding to the event E}}{\text{Total number of equally-likely outcomes}}$$

examples

If we roll a 6-sided die, calculate

1. P(rolling a 1)
2. P(rolling a number bigger than 4)

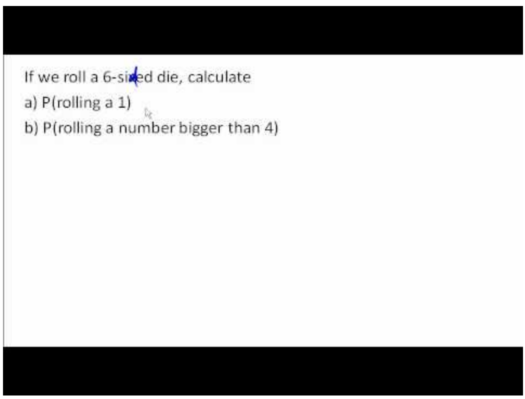
Solutions:

Recall that the sample space is {1,2,3,4,5,6}

1. There is one outcome corresponding to "rolling a 1," so the probability is $\frac{1}{6}$
2. There are two outcomes bigger than a 4, so the probability is $\frac{2}{6} = \frac{1}{3}$

Probabilities are essentially fractions, and can be reduced to lower terms like fractions.

This video describes this example and the previous one in detail.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=163

Let's say you have a bag with 20 cherries, 14 sweet and 6 sour. If you pick a cherry at random, what is the probability that it will be sweet?

Solution:

There are 20 possible cherries that could be picked, so the number of possible outcomes is 20. Of these 20 possible outcomes, 14 are favorable (sweet), so the probability that the cherry will be sweet is $\frac{14}{20} = \frac{7}{10}$.

There is one potential complication to this example, however. It must be assumed that the probability of picking any of the cherries is the same as the probability of picking any other. This wouldn't be true if (let us imagine) the sweet cherries are smaller than the sour ones. (The sour cherries would come to hand more readily when you sampled from the bag.) Let us keep in mind, therefore, that when we assess probabilities in terms of the ratio of favorable to all potential cases, we rely heavily on the assumption of equal probability for all outcomes.

Try It

At some random moment, you look at your clock and note the minutes reading.

a. What is probability the minutes reading is 15?

b. What is the probability the minutes reading is 15 or less?

Cards

A standard deck of 52 playing cards consists of four **suits** (hearts, spades, diamonds and clubs). Spades and clubs are black while hearts and diamonds are red. Each suit contains 13 cards, each of a different **rank**: an Ace (which in many games functions as both a low card and a high card), cards numbered 2 through 10, a Jack, a Queen and a King.

example

Compute the probability of randomly drawing one card from a deck and getting an Ace.

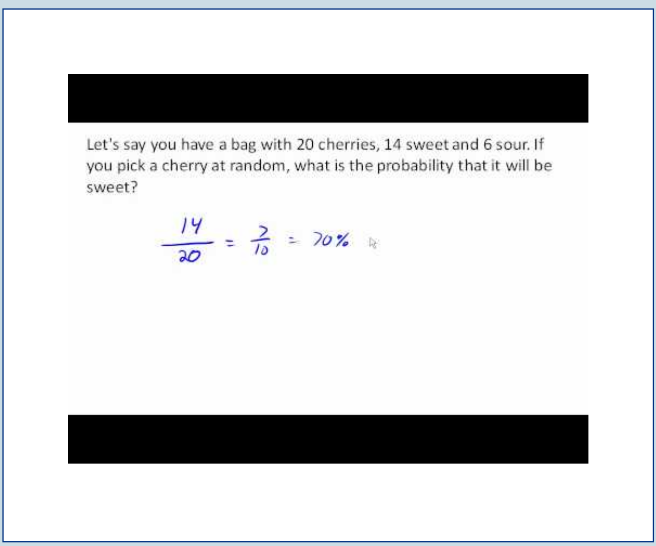
Solution:

There are 52 cards in the deck and 4 Aces so $P(Ace) = \frac{4}{52} = \frac{1}{13} \approx 0.0769$

We can also think of probabilities as percents: There is a 7.69% chance that a randomly selected card will be an Ace.

Notice that the smallest possible probability is 0 – if there are no outcomes that correspond with the event. The largest possible probability is 1 – if all possible outcomes correspond with the event.

This video demonstrates both this example and the previous cherry example on the page.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=163

Certain and Impossible events

- An impossible event has a probability of 0.
- A certain event has a probability of 1.
- The probability of any event must be $0 \leq P(E) \leq 1$

Try It

Click here to try this problem.

In the course of this section, **if you compute a probability and get an answer that is negative or greater than 1, you have made a mistake and should check your work**.

Types of Events

Complementary Events

Now let us examine the probability that an event does **not** happen. As in the previous section, consider the situation of rolling a six-sided die and first compute the probability of rolling a six: the answer is $P(\text{six}) = 1/6$. Now consider the probability that we do *not* roll a six: there are 5 outcomes that are not a six, so the answer is $P(\text{not a six}) = \frac{5}{6}$. Notice that

$$P(\text{six}) + P(\text{not a six}) = \frac{1}{6} + \frac{5}{6} = \frac{6}{6} = 1$$

This is not a coincidence. Consider a generic situation with n possible outcomes and an event E that corresponds to m of these outcomes. Then the remaining $n - m$ outcomes correspond to E not happening, thus

$$P(\text{not}E) = \frac{n-m}{n} = \frac{n}{n} - \frac{m}{n} = 1 - \frac{m}{n} = 1 - P(E)$$

Complement of an Event

The **complement** of an event is the event "*E* doesn't happen"

- The notation $\bar{E}$ is used for the complement of event *E*.
- We can compute the probability of the complement using $P\left(\bar{E}\right) = 1 - P(E)$
- Notice also that $P(E) = 1 - P\left(\bar{E}\right)$

example

If you pull a random card from a deck of playing cards, what is the probability it is not a heart?

Solution:

There are 13 hearts in the deck, so $P(\text{heart}) = \frac{13}{52} = \frac{1}{4}$.

The probability of *not* drawing a heart is the complement:

$P(\text{not heart}) = 1 - P(\text{heart}) = 1 - \frac{1}{4} = \frac{3}{4}$

This situation is explained in the following video.

https://youtu.be/RnljiW6ZM6A

Try It

Click here to try this problem.

Probability of two independent events

example

Suppose we flipped a coin and rolled a die, and wanted to know the probability of getting a head on the coin and a 6 on the die.

Solution:

We could list all possible outcomes: {H1,H2,h2,H4,H5,H6,T1,T2,T3,T4,T5,T6}.

Notice there are $2 \cdot 6 = 12$ total outcomes. Out of these, only 1 is the desired outcome, so the probability is $\frac{1}{12}$.

The prior example contained two *independent* events. Getting a certain outcome from rolling a die had no influence on the outcome from flipping the coin.

Independent Events

Events A and B are **independent events** if the probability of Event B occurring is the same whether or not Event A occurs.

example

Are these events independent?

1. A fair coin is tossed two times. The two events are (1) first toss is a head and (2) second toss is a head.
2. The two events (1) “It will rain tomorrow in Houston” and (2) “It will rain tomorrow in Galveston” (a city near Houston).
3. You draw a card from a deck, then draw a second card without replacing the first.

Solutions:

1. The probability that a head comes up on the

second toss is 1/2 regardless of whether or not a head came up on the first toss, so these events are independent.

2. These events are not independent because it is more likely that it will rain in Galveston on days it rains in Houston than on days it does not.
3. The probability of the second card being red depends on whether the first card is red or not, so these events are not independent.

When two events are independent, the probability of both occurring is the product of the probabilities of the individual events.

P(*A* and *B*) for independent events

If events *A* and *B* are independent, then the probability of both *A* and *B* occurring is

$$P\left(A\text{ and }B\right)=P\left(A\right)\cdot P\left(B\right)$$

where *P*(*A* and *B*) is the probability of events *A* and *B* both occurring, *P*(*A*) is the probability of event A occurring, and $P(B)$ is the probability of event B occurring

If you look back at the coin and die example from earlier, you can see how the number of outcomes of the first event multiplied by the number of outcomes in the second event multiplied to equal the total number of possible outcomes in the combined event.

example

In your drawer you have 10 pairs of socks, 6 of which are white, and 7 tee shirts, 3 of which are white. If you randomly reach in and pull out a pair of socks and a tee shirt, what is the probability both are white?

Solution:

The probability of choosing a white pair of socks is $\frac{6}{10}$.

The probability of choosing a white tee shirt is $\frac{3}{7}$.

The probability of both being white is $\frac{6}{10} \cdot \frac{3}{7} = \frac{18}{70} = \frac{9}{35}$

Examples of joint probabilities are discussed in this video.

https://youtu.be/6F17WLp-EL8

Try It

Click here to try this problem.

The previous examples looked at the probability of *both* events occurring. Now we will look at the probability of *either* event occurring.

example

Suppose we flipped a coin and rolled a die, and wanted to know the probability of getting a head on the coin *or* a 6 on the die.

Solution:

Here, there are still 12 possible outcomes: {H1,H2,h2,H4,H5,H6,T1,T2,T3,T4,T5,T6}

By simply counting, we can see that 7 of the outcomes have a head on the coin *or* a 6 on the die *or* both – we use *or* inclusively here (these 7 outcomes are H1, H2, h2, H4, H5, H6, T6), so the probability is $\frac{7}{12}$. How could we have found this from the individual probabilities?

As we would expect, $\frac{1}{2}$ of these outcomes have a head, and $\frac{1}{6}$ of these outcomes have a 6 on the die. If we add these, $\frac{1}{2}+\frac{1}{6}=\frac{6}{12}+\frac{2}{12}=\frac{8}{12}$, which is not the correct probability. Looking at the outcomes we can see why: the outcome H6 would have been counted twice, since it contains both a head and a 6; the probability of both a head *and* rolling a 6 is $\frac{1}{12}$.

If we subtract out this double count, we have the correct probability: $\frac{8}{12}-\frac{1}{12}=\frac{7}{12}$.

P(*A* or *B*)

The probability of either *A* or *B* occurring (or both) is

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

example

Suppose we draw one card from a standard deck. What is the probability that we get a Queen or a King?

Solution:

There are 4 Queens and 4 Kings in the deck, hence 8 outcomes corresponding to a Queen or King out of 52 possible outcomes. Thus the probability of drawing a Queen or a King is:

$$P(\text{King or Queen}) = \frac{8}{52}$$

Note that in this case, there are no cards that are both a Queen and a King, so

$P(\text{King and Queen}) = 0$. Using our probability rule, we could have said:

$P(\text{King or Queen}) = P(\text{King}) + P(\text{Queen}) - P(\text{King and Queen}) = \frac{4}{52} + \frac{4}{52} - 0 = \frac{8}{52}$

See more about this example and the previous one in the following video.

https://youtu.be/klbPZeH1np4

In the last example, the events were **mutually exclusive**, so $P(A \text{ or } B) = P(A) + P(B)$.

Try It

Click here to try this problem.

example

Suppose we draw one card from a standard deck.

What is the probability that we get a red card or a King?

Solution:

Half the cards are red, so $P(\text{red}) = \frac{26}{52}$

There are four kings, so $P(\text{King}) = \frac{4}{52}$

There are two red kings, so $P(\text{Red and King}) = \frac{2}{52}$

We can then calculate

$P(\text{Red or King}) = P(\text{Red}) + P(\text{King}) - P(\text{Red and King}) = \frac{26}{52} + \frac{4}{52} - \frac{2}{52} = \frac{28}{52}$

Try It

In your drawer you have 10 pairs of socks, 6 of which are white, and 7 tee shirts, 3 of which are white. If you reach in and randomly grab a pair of socks and a tee shirt, what the probability at least one is white?

Example

The table below shows the number of survey subjects who have received and not received a speeding ticket in the last year, and the color of their car. Find the probability that a randomly chosen person:

1. Has a red car *and* got a speeding ticket
2. Has a red car *or* got a speeding ticket.

	Speeding ticket	No speeding ticket	Total
Red car	15	135	150
Not red car	45	470	515
Total	60	605	665

Solution:

We can see that 15 people of the 665 surveyed had both a red car and got a speeding ticket, so the probability is $\frac{15}{665} \approx 0.0226$.

Notice that having a red car and getting a speeding ticket are not independent events, so the probability of both of them occurring is not simply the product of probabilities of each one occurring.

We could answer this question by simply adding up the numbers: 15 people with red cars and speeding tickets + 135 with red cars but no ticket + 45 with a ticket but no red car = 195 people. So the probability is $\frac{195}{665} \approx 0.2932$.

We also could have found this probability by:

P(had a red car) + P(got a speeding ticket) – P(had a red car and got a speeding ticket)

= $\frac{150}{665} + \frac{60}{665} - \frac{15}{665} = \frac{195}{665}$.

This table example is detailed in the following explanatory video.

https://youtu.be/HWrGoM1yRaU

Try It

Click here to try this problem.

Click here to try this problem.

Conditional Probability

In the previous section we computed the probabilities of events that were independent of each other. We saw that getting a certain outcome from rolling a die had no influence on the outcome from flipping a coin, even though we were computing a probability based on doing them at the same time.

In this section, we will consider events that *are* dependent on each other, called **conditional probabilities**.

Conditional Probability

The probability the event B occurs, given that event A has happened, is represented as

$$P(B \mid A)$$

This is read as "the probability of B given A"

For example, if you draw a card from a deck, then the sample space for the next card drawn has changed, because you are now working with a deck of 51 cards. In the following example we will show you how the computations for events like this are different from the computations we did in the last section.

example

What is the probability that two cards drawn at random from a deck of playing cards will both be aces?

Solution:

It might seem that you could use the formula for the probability of two independent events and simply multiply $\frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}$. This would be incorrect, however, because the two events are not independent. If the first card drawn is an ace, then the probability that the second card is also an ace would be lower because there would only be three aces left in the deck.

Once the first card chosen is an ace, the probability that the second card chosen is also an ace is called the **conditional probability** of drawing an ace. In this case the "condition" is that the first card is an ace. Symbolically, we write this as:

P(ace on second draw | an ace on the first draw).

The vertical bar "|" is read as "given," so the above expression is short for "The probability that an ace is drawn on the second draw given that an ace was drawn on the first draw." What is this probability? After an ace is drawn on the first draw, there are 3 aces out of 51 total cards left. This means that the conditional probability of drawing an ace after one ace has already been drawn is $\frac{3}{51} = \frac{1}{17}$.

Thus, the probability of both cards being aces is $\frac{4}{52} \cdot \frac{3}{51} = \frac{12}{2652} = \frac{1}{221}$.

Conditional Probability Formula

If Events A and B are not independent, then

$P(A \text{ and } B) = P(A) \cdot P(B \mid A)$

example

If you pull 2 cards out of a deck, what is the probability that both are spades?

Solution:

The probability that the first card is a spade is $\frac{13}{52}$.

The probability that the second card is a spade, given the first was a spade, is $\frac{12}{51}$, since there is one less spade in the deck, and one less total cards.

The probability that both cards are spades is $\frac{13}{52} \cdot \frac{12}{51} = \frac{156}{2652} \approx 0.0588$

Try It

Click here to try this problem.

Example

The table below shows the number of survey subjects who have received and not received a speeding ticket in the last year, and the color of their car. Find the probability that a randomly chosen person:

1. has a speeding ticket *given* they have a red car
2. has a red car *given* they have a speeding ticket

Speeding ticket	No speeding ticket	Total	
Red car	15	135	150
Not red car	45	470	515
Total	60	605	665

Solutions:

1. Since we know the person has a red car, we are only considering the 150 people in the first row of the table. Of those, 15 have a speeding ticket, so P(ticket | red car) = $\frac{15}{150} = \frac{1}{10} = 0.1$
2. Since we know the person has a speeding ticket, we are only considering the 60 people in the first column of the table. Of those, 15 have a red car, so P(red car | ticket) = $\frac{15}{60} = \frac{1}{4} = 0.25$.

Notice from the last example that P(B | A) is **not** equal to P(A | B).

These kinds of conditional probabilities are what insurance companies use to determine your insurance rates. They look at the conditional probability of you having accident, given your age, your car, your car color, your driving history, etc., and price your policy based on that likelihood.

View more about conditional probability in the following video.

https://youtu.be/b6tstekMlb8

Example

If you draw two cards from a deck, what is the probability that you will get the Ace of Diamonds and a black card?

Solution:

You can satisfy this condition by having Case A or Case B, as follows:

Case A) you can get the Ace of Diamonds first and then a black card or

Case B) you can get a black card first and then the Ace of Diamonds.

Let's calculate the probability of Case A. The probability that the first card is the Ace of Diamonds is $\frac{1}{52}$. The probability that the second card is black given that the first card is the Ace of Diamonds is $\frac{26}{51}$ because 26 of the remaining 51 cards are black. The probability is therefore $\frac{1}{52} \cdot \frac{26}{51} = \frac{1}{102}$.

Now for Case B: the probability that the first card is black is $\frac{26}{52} = \frac{1}{2}$. The probability that the second card is the Ace of Diamonds given that the first card is black is $\frac{1}{51}$. The probability of Case B is therefore $\frac{1}{2} \cdot \frac{1}{51} = \frac{1}{102}$, the same as the probability of Case 1.

Recall that the probability of A or B is *P*(A) + *P*(B) – *P*(A and B). In this problem, *P*(A and B) = 0 since the first card cannot be the Ace of Diamonds and be a black card. Therefore, the probability of Case A or Case B is $\frac{1}{101} + \frac{1}{101} = \frac{2}{101}$. The probability that you will get the Ace of Diamonds and a black card when drawing two cards from a deck is $\frac{2}{101}$.

These two playing card scenarios are discussed further in the following video.

[Click here to view this video.]

Try It

Click here to try this problem.

Example

A home pregnancy test was given to women, then pregnancy was verified through blood tests. The following table shows the home pregnancy test results.

Find

1. P (not pregnant | positive test result)
2. P (positive test result | not pregnant)

	Positive test	Negative test	Total
Pregnant	70	4	74
Not Pregnant	5	14	19
Total	75	18	93

Solutions:

1. Since we know the test result was positive, we're limited to the 75 women in the first column, of which 5 were not pregnant. *P*(not pregnant | positive test result) = $\frac{5}{75} \approx 0.067$.
2. Since we know the woman is not pregnant, we are limited to the 19 women in the second row, of which 5 had a positive test. *P*(positive test result | not pregnant) = $\frac{5}{19} \approx 0.263$

The second result is what is usually called a false positive: A positive result when the woman is not actually pregnant.

See more about this example here.

Click here to view this video.

Try It

Click here to try this problem.

Attributions

This chapter contains material taken from *Math in Society*

APPLICATIONS WITH PROBABILITY

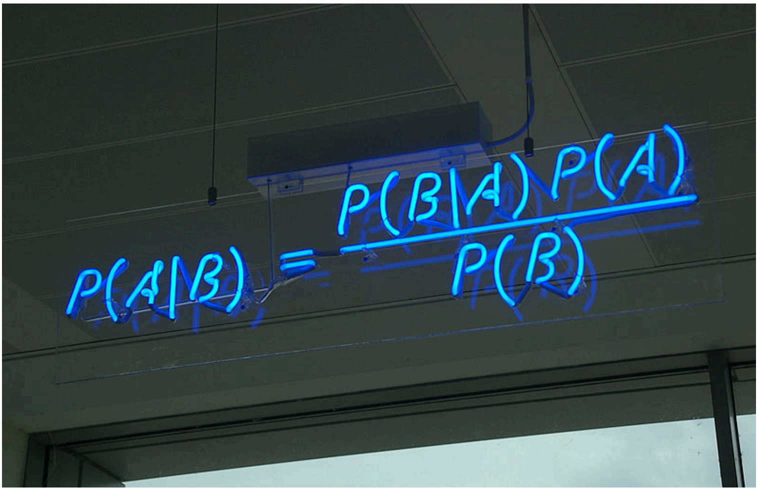
Learning Outcomes

- Compute a conditional probability for an event
- Use Baye's theorem to compute a conditional probability
- Calculate the expected value of an event

In the next section, we will explore more complex conditional probabilities and ways to compute them. Conditional probabilities can give us information such as the likelihood of getting a positive test result for a disease without actually having the disease. If a doctor thinks the chances that a positive test result nearly guarantees that a patient has a disease, they might begin an unnecessary and possibly harmful treatment regimen on a healthy patient. If you were to get a positive test result, knowing the likelihood of getting a false positive can guide you to get a second opinion.

Bayes' Theorem

In this section we concentrate on the more complex conditional probability problems we began looking at in the last section.



For example, suppose a certain disease has an incidence rate of 0.1% (that is, it afflicts 0.1% of the population). A test has been devised to detect this disease. The test does not produce false negatives (that is, anyone who has the disease will test positive for it), but the false positive rate is 5% (that is, about 5% of people who take the test will test positive, even though they do not have the disease). Suppose a randomly selected person takes the test and tests positive. What is the probability that this person actually has the disease?

There are two ways to approach the solution to this problem. One involves an important result in probability theory called Bayes' theorem. We will discuss this theorem a bit later, but for now we will use an alternative and, we hope, much more intuitive approach.

Let's break down the information in the problem piece by piece as an example.

example

Suppose a certain disease has an incidence rate of 0.1% (that is, it afflicts 0.1% of the population). The percentage 0.1% can be converted to a decimal number by moving the decimal place two places to the left, to get 0.001. In turn, 0.001 can be rewritten as a fraction: 1/1000. This tells us that about 1 in every 1000 people has the disease. (If we wanted we could write $P(\text{disease})=0.001$.)

A test has been devised to detect this disease. The test does not produce false negatives (that is, anyone who has the disease will test positive for it). This part is fairly straightforward: everyone who has the disease will test positive, or alternatively everyone who tests negative does not have the disease. (We could also say $P(\text{positive} \mid \text{disease})=1$.)

The false positive rate is 5% (that is, about 5%

of people who take the test will test positive, even though they do not have the disease). This is even more straightforward. Another way of looking at it is that of every 100 people who are tested and do not have the disease, 5 will test positive even though they do not have the disease. (We could also say that P(positive | no disease)=0.05.)

Suppose a randomly selected person takes the test and tests positive. What is the probability that this person actually has the disease? Here we want to compute P(disease|positive). We already know that P(positive|disease)=1, but remember that conditional probabilities are not equal if the conditions are switched.

Rather than thinking in terms of all these probabilities we have developed, let's create a hypothetical situation and apply the facts as set out above. First, suppose we randomly select 1000 people and administer the test. How many do we expect to have the disease? Since about 1/1000 of all people are afflicted with the disease, 1/1000 of 1000 people is 1. (Now you know why we chose 1000.) Only 1 of 1000 test subjects actually has the disease; the other 999 do not.

We also know that 5% of all people who do not have the disease will test positive. There are 999 disease-free people, so we would expect (0.05)(999)=49.95 (so, about 50) people to test positive who do not have the disease.

Now back to the original question, computing *P*(disease|positive). There are 51 people who test positive in our example (the one unfortunate person who actually has the disease, plus the 50 people who tested positive but don't). Only one of these people has the disease, so

$$\text{P(disease | positive)} \approx \frac{1}{51} \approx 0.0196$$

or less than 2%. Does this surprise you? This means that of all people who test positive, over 98% *do not have the disease*.

The answer we got was slightly approximate, since we rounded 49.95 to 50. We could redo the problem with 100,000 test subjects, 100 of whom would have the disease and (0.05)(99,900)=4995 test positive but do not have the disease, so the exact probability of having the disease if you test positive is

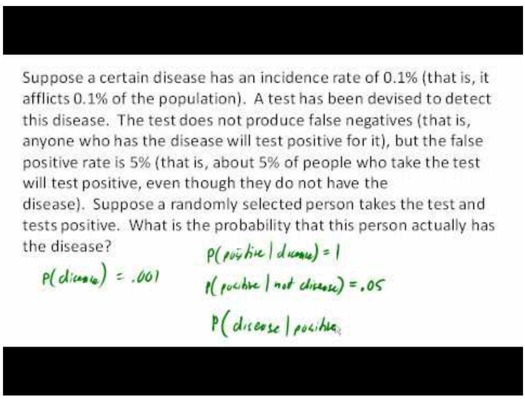
$$\text{P(disease | positive)} \approx \frac{100}{5095} \approx 0.0196$$

which is pretty much the same answer.

But back to the surprising result. *Of all people who* *test positive, over 98% do not have the disease.* If your guess for the probability a person who tests positive has the disease was wildly different from the right answer (2%), don't feel bad. The exact same problem was posed to doctors and medical students at the Harvard Medical School 25 years ago and the results revealed in a 1978 *New England Journal of Medicine* article. Only about 18% of the participants got the right answer. Most of the rest thought the answer was closer to 95% (perhaps they were misled by the false positive rate of 5%).

So at least you should feel a little better that a bunch of doctors didn't get the right answer either (assuming you thought the answer was much higher). But the significance of this finding and similar results from other studies in the intervening years lies not in making math students feel better but in the possibly catastrophic consequences it might have for patient care. If a doctor thinks the chances that a positive test result nearly guarantees that a patient has a disease, they might begin an unnecessary and possibly harmful treatment regimen on a healthy patient. Or worse, as in the early days of the AIDS crisis when being HIV-positive was often equated with a death sentence, the patient might take a drastic action and commit suicide.

This example is worked through in detail in the video here.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

As we have seen in this hypothetical example, the most responsible course of action for treating a patient who tests positive would be to counsel the patient that they most likely do *not* have the disease and to order further, more reliable, tests to verify the diagnosis.

One of the reasons that the doctors and medical students in the study did so poorly is that such problems, when presented in the types of statistics courses that medical students often take, are solved by use of Bayes' theorem, which is stated as follows:

Bayes' Theorem

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(\bar{A})P(B|\bar{A})}$$

In our earlier example, this translates to

$$P(\text{disease}|\text{positive}) = \frac{P(\text{disease})P(\text{positive}|\text{disease})}{P(\text{disease})P(\text{positive}|\text{disease})+P(\text{nodisease})P(\text{positive}|\text{nodisease})}$$

Plugging in the numbers gives

$$P(\text{disease}|\text{positive}) = \frac{(0.001)(1)}{(0.001)(1)+(0.999)(0.05)} \approx 0.0196$$

which is exactly the same answer as our original solution.

The problem is that you (or the typical medical student, or even the typical math professor) are much more likely to be able to remember the original solution than to remember Bayes' theorem. Psychologists, such as Gerd Gigerenzer, author of *Calculated Risks: How to Know When Numbers Deceive You*, have advocated that the method involved in the original solution (which Gigerenzer calls the method of "natural frequencies") be employed in place of Bayes'

Theorem. Gigerenzer performed a study and found that those educated in the natural frequency method were able to recall it far longer than those who were taught Bayes' theorem. When one considers the possible life-and-death consequences associated with such calculations it seems wise to heed his advice.

example

A certain disease has an incidence rate of 2%. If the false negative rate is 10% and the false positive rate is 1%, compute the probability that a person who tests positive actually has the disease.

Solution:

Imagine 10,000 people who are tested. Of these 10,000, 200 will have the disease; 10% of them, or 20, will test negative and the remaining 180 will test positive. Of the 9800 who do not have the disease, 98 will test positive. So of the 278 total people who test positive, 180 will have the disease. Thus

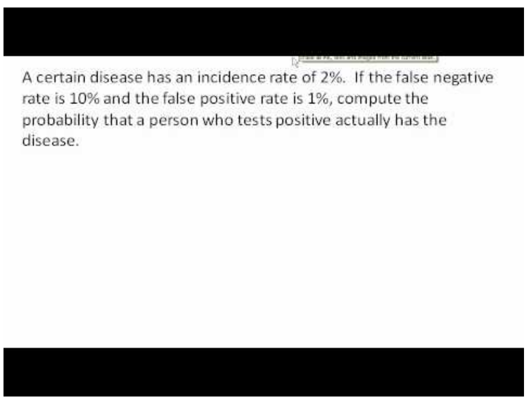
$$P(\text{disease}|\text{positive}) = \frac{180}{278} \approx 0.647$$

so about 65% of the people who test positive will have the disease.

Using Bayes theorem directly would give the same result:

$$P(\text{disease}|\text{positive}) = \frac{(0.02)(0.90)}{(0.02)(0.90)+(0.98)(0.01)} = \frac{0.018}{0.0278} \approx 0.647$$

View the following for more about this example.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

Try It

Click here to try this problem.

Counting

Counting? You already know how to count or you wouldn't be taking a college-level math class, right? Well yes, but what we'll really be investigating here are ways of counting *efficiently*. When we get to the probability situations a bit later in this chapter we will need to count some *very* large numbers, like the number of possible winning lottery tickets. One way to do this would be to write down every possible set of numbers that might show up on a lottery ticket, but believe me: you don't want to do this.

Basic Counting

We will start, however, with some more reasonable sorts of counting problems in order to develop the ideas that we will soon need.

example

Suppose at a particular restaurant you have three choices for an appetizer (soup, salad or breadsticks) and five choices for a main course (hamburger, sandwich, quiche, fajita or pizza). If you are allowed to choose exactly one item from each category for your meal, how many different meal options do you have?

Solution 1: One way to solve this problem would be to systematically list each possible meal:

soup + hamburger soup + sandwich soup + quiche

soup + fajita soup + pizza salad + hamburger

salad + sandwich salad + quiche salad + fajita

salad + pizza breadsticks + hamburger breadsticks + sandwich

breadsticks + quiche breadsticks + fajita breadsticks + pizza

Assuming that we did this systematically and that we neither missed any possibilities nor listed any possibility more than once, the answer would be 15. Thus you could go to the restaurant 15 nights in a row and have a different meal each night.

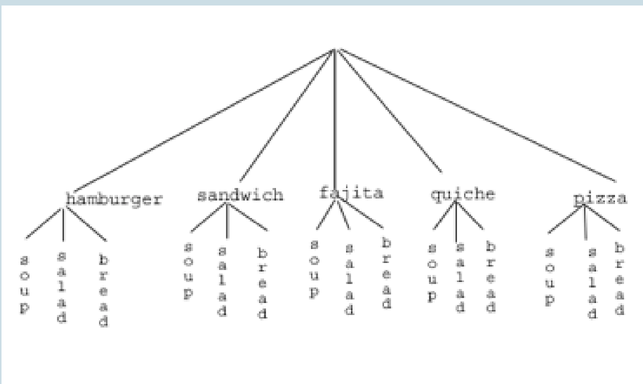
Solution 2: Another way to solve this problem would be to list all the possibilities in a table:

	hamburger	sandwich	quiche	fajita	pizza
soup	soup+burger				
salad	salad+burger				
bread	etc				

In each of the cells in the table we could list the corresponding meal: soup + hamburger in the upper left corner, salad + hamburger below it, etc. But if we didn't really care *what* the possible meals are, only *how many* possible meals there are, we could just count the number of cells and arrive at an answer of 15, which matches our answer from the first solution. (It's always good when you solve a problem two different ways and get the same answer!)

Solution 3: We already have two perfectly good solutions. Why do we need a third? The first method was not very systematic, and we might easily have made an omission. The second method was better, but suppose that in addition to the appetizer and the main course we further complicated the problem by adding desserts to the menu: we've used the rows of the table for the appetizers and the columns for the main courses—where will the desserts go? We would need a third dimension, and since drawing 3-D tables on a 2-D page or computer screen isn't terribly easy, we need a better way in case we have three categories to choose form instead of just two.

So, back to the problem in the example. What else can we do? Let's draw a **tree diagram**:



This is called a "tree" diagram because at each stage we branch out, like the branches on a tree. In this case, we first drew five branches (one for each main course) and then for each of those branches we drew three more branches (one for each appetizer). We count the number of branches at the final level and get (surprise, surprise!) 15.

If we wanted, we could instead draw three branches at the first stage for the three appetizers and then five branches (one for each main course) branching out of each of those three branches.

OK, so now we know how to count possibilities using tables and tree diagrams. These methods will continue to be useful in certain cases, but imagine a game where you have two decks of cards (with 52 cards in each deck) and you select one card from each deck. Would you really want to draw a table or tree diagram to determine the number of outcomes of this game?

Let's go back to the previous example that involved selecting a meal from three appetizers and five main courses, and look at the second solution that used a table. Notice that one way to count the number of possible meals is simply to number each of the appropriate cells in the table, as we have done above. But another way to count the number of cells in the table would be multiply the number of rows (3) by the number of columns (5) to get 15. Notice that we could have arrived at the same result without making a table at all by simply multiplying the number of choices for the appetizer (3) by the number of choices for the main course (5). We generalize this technique as the *basic counting rule*:

Basic Counting Rule

If we are asked to choose one item from each of two separate categories where there are m items in the first category and n items in the second category, then the total number of available choices is $m \cdot n$.

This is sometimes called the multiplication rule for probabilities.

example

There are 21 novels and 18 volumes of poetry on a reading list for a college English course. How many different ways can a student select one novel and one volume of poetry to read during the quarter?

Solution:

There are 21 choices from the first category and 18 for the second, so there are $21 \cdot 18 = 378$ possibilities.

The Basic Counting Rule can be extended when there are more than two categories by applying it repeatedly, as we see in the next example.

example

Suppose at a particular restaurant you have three choices for an appetizer (soup, salad or breadsticks), five choices for a main course (hamburger, sandwich, quiche, fajita or pasta) and two choices for dessert (pie or ice cream). If you are allowed to choose exactly one item from each category for your meal, how many different meal options do you have?

Solution:

There are 3 choices for an appetizer, 5 for the main course and 2 for dessert, so there are $3 \cdot 5 \cdot 2 = 30$ possibilities.

Try It

Click here to try this problem.

Example

A quiz consists of 3 true-or-false questions. In how many ways can a student answer the quiz?

Solution:

There are 3 questions. Each question has 2 possible answers (true or false), so the quiz may be answered in $2 \cdot 2 \cdot 2 = 8$ different ways. Recall that another way to write $2 \cdot 2 \cdot 2$ is 2^3, which is much more compact.

Basic counting examples from this section are described in the following video.

[Click here to view this video.]

Permutations

In this section we will develop an even faster way to solve some of the problems we have already learned to solve by other means. Let's start with a couple examples.

example

How many different ways can the letters of the word MATH be rearranged to form a four-letter code word?

Solution:
This problem is a bit different. Instead of choosing one item from each of several different categories, we are repeatedly choosing items from the *same* category (the category is: the letters of the word MATH) and each time we choose an item we *do not replace* it, so there is one fewer choice at the next stage: we have 4 choices for the first letter (say we choose A), then 3 choices for the second (M, T and H; say we choose H), then 2 choices for the next letter (M and T; say we choose M) and only one choice at the last stage (T). Thus there are $4 \bullet 3 \bullet 2 \bullet 1 = 24$ ways to spell a code worth with the letters MATH.

In this example, we needed to calculate $n \cdot (n - 1) \cdot (n - 2) \cdots 3 \cdot 2 \cdot 1$. This calculation shows up often in mathematics, and is called the **factorial**, and is notated $n!$

Factorial

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$$

Try It

Click here to try this problem.

example

How many ways can five different door prizes be distributed among five people?

Solution:

There are 5 choices of prize for the first person, 4 choices for the second, and so on. The number of ways the prizes can be distributed will be $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$ ways.

Try It

Click here to try this problem.

Now we will consider some slightly different examples.

examples

A charity benefit is attended by 25 people and three gift certificates are given away as door prizes: one gift certificate is in the amount of $100, the second is worth $25 and the third is worth $10. Assuming that no person receives more than one prize, how many different ways can the three gift certificates be awarded?

Solution:

Using the Basic Counting Rule, there are 25 choices for the person who receives the $100 certificate, 24 remaining choices for the $25 certificate and 23 choices for the $10 certificate, so there are 25 • 24 • 23 = 13,800 ways in which the prizes can be awarded.

Example

Eight sprinters have made it to the Olympic finals in the 100-meter race. In how many different ways can the gold, silver and bronze medals be awarded?

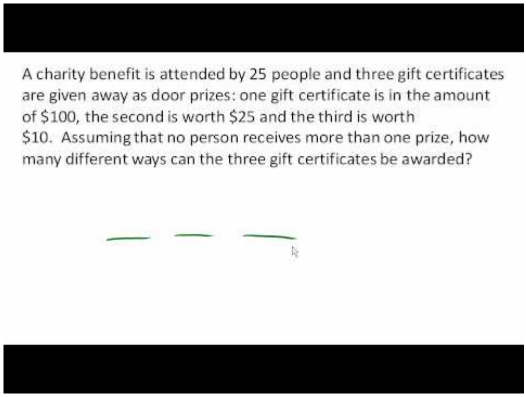
Solution:

Using the Basic Counting Rule, there are 8 choices for the gold medal winner, 7 remaining choices for the silver, and 6 for the bronze, so there are 8 • 7 • 6 = 336 ways the three medals can be awarded to the 8 runners.

Note that in these preceding examples, the gift certificates and the Olympic medals were awarded *without replacement*; that is, once we have chosen a winner of the first door prize or the gold medal, they are not eligible for the other prizes. Thus, at each succeeding stage of the solution there is one fewer choice (25, then 24, then 23 in the first example; 8, then 7, then 6 in the second). Contrast this with the situation of a multiple choice test, where there might be five possible answers — A, B, C, D or E — for each question on the test.

Note also that *the order of selection was important* in each example: for the three door prizes, being chosen first means that you receive substantially more money; in the Olympics example, coming in first means that you get the gold medal instead of the silver or bronze. In each case, if we had chosen the same three people in a different order there might have been a different person who received the $100 prize, or a different goldmedalist. (Contrast this with the situation where we might draw three names out of a hat to each receive a $10 gift certificate; in this case the order of selection is *not* important since each of the three people receive the same prize. Situations where the order is *not* important will be discussed in the next section.)

Factorial examples are worked in this video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

We can generalize the situation in the two examples above to any problem *without replacement* where the *order of selection is important*. If we are arranging in order r items out of n possibilities (instead of 3 out of 25 or 3 out of 8 as in the previous examples), the number of possible arrangements will be given by

$$n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)$$

If you don't see why $(n - r + 1)$ is the right number to use for the last factor, just think back to the first example in this section, where we calculated 25 • 24 • 23 to get 13,800. In this case $n = 25$ and $r = 3$, so $n - r + 1 = 25 - 3 + 1 = 23$, which is exactly the right number for the final factor.

Now, why would we want to use this complicated formula when it's actually easier to use the Basic Counting Rule, as we did in the first two examples? Well, we won't actually use this formula all that often; we only developed it so that we could attach a special notation and a special definition to this situation where we are choosing r items out of n possibilities *without replacement* and where the *order of selection is important*. In this situation we write:

Permutations

$$nPr = n \cdot (n-1) \cdot (n-2) \cdots (n-r+1)$$

We say that there are nPr **permutations** of size r that may be selected from among n choices *without replacement* when *order matters*.

It turns out that we can express this result more simply using factorials.

$$_nP_r = \frac{n!}{(n-r)!}$$

In practicality, we usually use technology rather than factorials or repeated multiplication to compute permutations.

example

I have nine paintings and have room to display only four of them at a time on my wall. How many different ways could I do this?

Solution:

Since we are choosing 4 paintings out of 9 *without replacement* where the *order of selection is important* there are $9P4 = 9 \cdot 8 \cdot 7 \cdot 6 = 3{,}024$ permutations.[/hidden-answer]

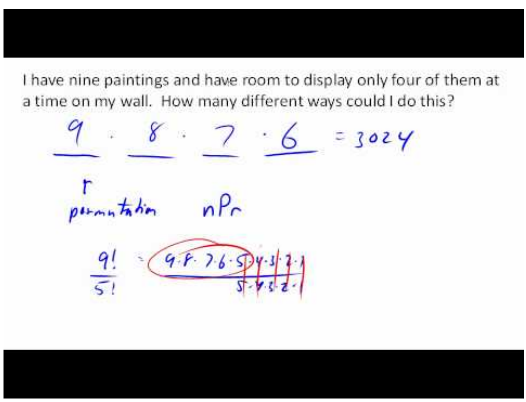
Example

How many ways can a four-person executive committee (president, vice-president, secretary, treasurer) be selected from a 16-member board of directors of a non-profit organization?

Solution:

We want to choose 4 people out of 16 without replacement and where the order of selection is important. So the answer is $16P4 = 16 \cdot 15 \cdot 14 \cdot 13 = 43{,}680$.[/hidden-answer]

View this video to see more about the permutations examples.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

Try It

How many 5 character passwords can be made using the letters A through Z

- if repeats are allowed
- if no repeats are allowed

Click here to try this problem.

Combinations

In the previous section we considered the situation where we chose r items out of n possibilities *without replacement* and where the *order of selection was important*. We now consider a similar situation in which the order of selection is *not* important.

Example

A charity benefit is attended by 25 people at which three $50 gift certificates are given away as door prizes. Assuming no person receives more than one prize, how many different ways can the gift certificates be awarded?

Solution:

Using the Basic Counting Rule, there are 25 choices for the first person, 24 remaining choices for the second person and 23 for the third, so there are $25 \cdot 24 \cdot 23 = 13{,}800$ ways to choose three people. Suppose for a moment that Abe is chosen first, Bea second and Cindy third; this is one of the 13,800 possible outcomes. Another way to award the prizes would be to choose Abe first, Cindy second and Bea third; this is another of the 13,800 possible outcomes. But either way Abe, Bea and Cindy each get $50, so it doesn't really matter the order in which we select them. In how many different orders can Abe, Bea and Cindy be selected? It turns out there are 6:

ABC ACB BAC BCA CAB CBA

How can we be sure that we have counted them all? We are really just choosing 3 people out of 3, so there are $3 \cdot 2 \cdot 1 = 6$ ways to do this; we didn't really need to list them all. We can just use permutations!

So, out of the 13,800 ways to select 3 people out of

25, six of them involve Abe, Bea and Cindy. The same argument works for any other group of three people (say Abe, Bea and David or Frank, Gloria and Hildy) so each three-person group is counted *six times*. Thus the 13,800 figure is six times too big. The number of distinct three-person groups will be 13,800/6 = 2300.

We can generalize the situation in this example above to any problem of choosing a collection of items *without replacement* where the *order of selection is **not** important*. If we are choosing r items out of n possibilities (instead of 3 out of 25 as in the previous examples), the number of possible choices will be given by $\frac{{}_nP_r}{{}_rP_r}$, and we could use this formula for computation. However this situation arises so frequently that we attach a special notation and a special definition to this situation where we are choosing r items out of n possibilities *without replacement* where the *order of selection is **not** important*.

Combinations

$$_nC_r = \frac{_nP_r}{_rP_r}$$

We say that there are *nCr* **combinations** of size *r* that may be selected from among *n* choices *without replacement* where *order doesn't matter*.

We can also write the combinations formula in terms of factorials:

$$_nC_r = \frac{n!}{(n-r)!r!}$$

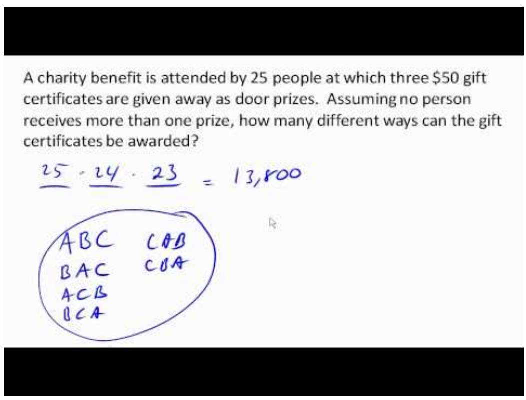
Example

A group of four students is to be chosen from a 35-member class to represent the class on the student council. How many ways can this be done?

Solution:

Since we are choosing 4 people out of 35 *without replacement* where the *order of selection is* ***not*** *important* there are ${}_{35}C_4 = \frac{35 \cdot 34 \cdot 33 \cdot 32}{4 \cdot 3 \cdot 2 \cdot 1}$ = 52,360 combinations.

View the following for more explanation of the combinations examples.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

Try It

The United States Senate Appropriations Committee consists of 29 members; the Defense Subcommittee of the Appropriations Committee consists of 19 members. Disregarding party affiliation or any special seats on the Subcommittee, how many different 19-member subcommittees may be chosen from among the 29 Senators on the Appropriations Committee?

In the preceding Try It problem we assumed that the 19 members of the Defense Subcommittee were chosen without regard to party affiliation. In reality this would never happen: if Republicans are in the majority they would never let a majority of Democrats sit on (and thus control) any subcommittee. (The same of course would be true if the Democrats were in control.) So let's consider the problem again, in a slightly more complicated form:

Example

The United States Senate Appropriations Committee consists of 29 members, 15 Republicans and 14 Democrats. The Defense Subcommittee consists of 19 members, 10 Republicans and 9 Democrats. How many different ways can the members of the Defense Subcommittee be chosen from among the 29 Senators on the Appropriations Committee?

Solution:

In this case we need to choose 10 of the 15 Republicans and 9 of the 14 Democrats. There are $15C10 = 3003$ ways to choose the 10 Republicans and $14C9 = 2002$ ways to choose the 9 Democrats. But now what? How do we finish the problem?

Suppose we listed all of the possible 10-member Republican groups on 3003 slips of red paper and all of the possible 9-member Democratic groups on 2002 slips of blue paper. How many ways can we choose one red slip and one blue slip? This is a job for the Basic Counting Rule! We are simply making one choice from the first category and one choice from the second category, just like in the restaurant menu problems from earlier.

There must be 3003 · 2002 = 6,012,006 possible ways of selecting the members of the Defense Subcommittee.

This example is worked through below.

A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

Probability Using Permutations and Combinations

We can use permutations and combinations to help us answer more complex probability questions.

examples

A 4 digit PIN number is selected. What is the probability that there are no repeated digits?

Solution:

There are 10 possible values for each digit of the PIN (namely: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9), so there are $10 \cdot 10 \cdot 10 \cdot 10 = 10^4 = 10000$ total possible PIN numbers.

To have no repeated digits, all four digits would have to be different, which is selecting without replacement. We could either compute $10 \cdot 9 \cdot 8 \cdot 7$, or notice that this is the same as the permutation $10P4 = 5040$.

The probability of no repeated digits is the number of 4 digit PIN numbers with no repeated digits divided by the total number of 4 digit PIN numbers. This probability is $\frac{{}_{10}P_4}{10^4} = \frac{5040}{10000} = 0.504$

Try It

Click here to try this problem.

Example

In a certain state's lottery, 48 balls numbered 1 through 48 are placed in a machine and six of them are drawn at random. If the six numbers drawn match the numbers that a player had chosen, the player wins $1,000,000. In this lottery, the order the numbers are drawn in doesn't matter. Compute the probability that you win the million-dollar prize if you purchase a single lottery ticket.

Solution:

In order to compute the probability, we need to count the total number of ways six numbers can be drawn, and the number of ways the six numbers on the player's ticket could match the six numbers drawn from the machine. Since there is no stipulation that the numbers be in any particular order, the number of possible outcomes of the lottery drawing is $48C6$ = 12,271,512. Of these possible outcomes, only one would match all six numbers on the player's ticket, so the probability of winning the grand prize is:

$$\frac{{}_6C_6}{{}_{48}C_6} = \frac{1}{12271512} \approx 0.0000000815$$

Example

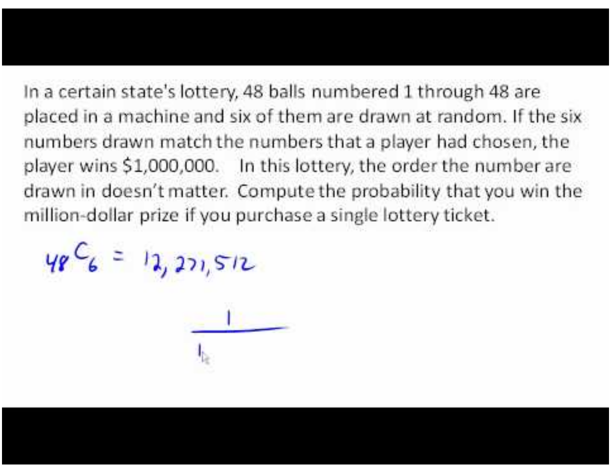
In the state lottery from the previous example, if five of the six numbers drawn match the numbers that a player has chosen, the player wins a second prize of $1,000. Compute the probability that you win the second prize if you purchase a single lottery ticket.

Solution:

As above, the number of possible outcomes of the lottery drawing is $48C6$ = 12,271,512. In order to win the second prize, five of the six numbers on the ticket must match five of the six winning numbers; in other words, we must have chosen five of the six winning numbers and one of the 42 losing numbers. The number of ways to choose 5 out of the 6 winning numbers is given by $6C5 = 6$ and the number of ways to choose 1 out of the 42 losing numbers is given by $42C1 = 42$. Thus the number of favorable outcomes is then given by the Basic Counting Rule: $6C5 \cdot 42C1 = 6 \cdot 42 = 252$. So the probability of winning the second prize is.

$$\frac{({}_6C_5)({}_{42}C_1)}{{}_{48}C_6} = \frac{252}{12271512} \approx 0.0000205$$

The previous examples are worked in the following video.



A YouTube element has been excluded from this

version of the text. You can view it online here: *https://granite.pressbooks.pub/math502/?p=165*

examples

Compute the probability of randomly drawing five cards from a deck and getting exactly one Ace.

Solution:

In many card games (such as poker) the order in which the cards are drawn is not important (since the player may rearrange the cards in his hand any way he chooses); in the problems that follow, we will assume that this is the case unless otherwise stated. Thus we use combinations to compute the possible number of 5-card hands, $52C5$. This number will go in the denominator of our probability formula, since it is the number of possible outcomes.

For the numerator, we need the number of ways to draw one Ace and four other cards (none of them Aces) from the deck. Since there are four Aces and we want exactly one of them, there will be $4C1$ ways to select one Ace; since there are 48 non-Aces and we want 4 of them, there will be $48C4$ ways to select the four non-Aces. Now we use the Basic Counting Rule to calculate that there will be $4C1 \cdot 48C4$ ways to choose one ace and four non-Aces.

Putting this all together, we have

$$P(\text{oneAce}) = \frac{({}_4C_1)({}_{48}C_4)}{{}_{52}C_5} = \frac{778320}{2598960} \approx 0.299$$

Example

Compute the probability of randomly drawing five cards from a deck and getting exactly two Aces.

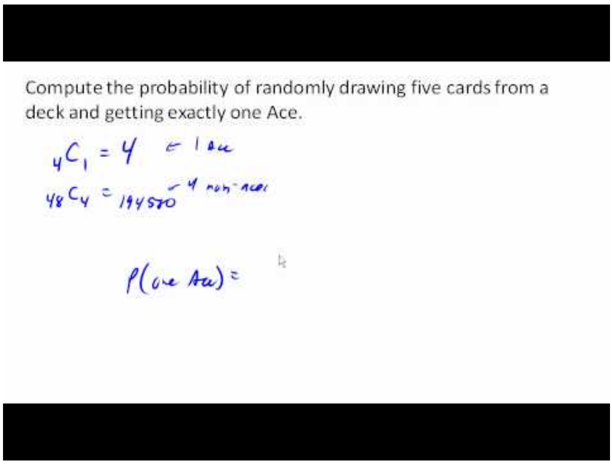
Solution:

The solution is similar to the previous example, except now we are choosing 2 Aces out of 4 and 3 non-Aces out of 48; the denominator remains the same:

$$P(\text{twoAces}) = \frac{({}_4C_2)({}_{48}C_3)}{{}_{52}C_5} = \frac{103776}{2598960} \approx 0.0399$$

It is useful to note that these card problems are remarkably similar to the lottery problems discussed earlier.

View the following for further demonstration of these examples.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

Try It

[Click here to try this problem.]()

Birthday Problem

Let's take a pause to consider a famous problem in probability theory:

> Suppose you have a room full of 30 people. What is the probability that there is at least one shared birthday?

Take a guess at the answer to the above problem. Was your guess fairly low, like around 10%? That seems to be the intuitive answer (30/365, perhaps?). Let's see if we should listen to our intuition. Let's start with a simpler problem, however.

example

Suppose three people are in a room. What is the probability that there is at least one shared birthday among these three people?

Solution:

There are a lot of ways there could be at least one shared birthday. Fortunately there is an easier way. We ask ourselves “What is the alternative to having at least one shared birthday?” In this case, the alternative is that there are **no** shared birthdays. In other words, the alternative to “at least one” is having **none**. In other words, since this is a complementary event,

$$P(\text{at least one}) = 1 - P(\text{none})$$

We will start, then, by computing the probability that there is no shared birthday. Let’s imagine that you are one of these three people. Your birthday can be anything without conflict, so there are 365 choices out of 365 for your birthday. What is the probability that the second person does not share your birthday? There are 365 days in the year (let’s ignore leap years) and removing your birthday from contention, there are 364 choices that will guarantee that you do not share a birthday with this person, so the probability that the second person does not share your birthday is 364/365. Now we move to the third person. What is the probability that this third person does not have the same birthday as either you or the second person? There are 363 days that will not duplicate your birthday or the second person’s, so the probability that the third person does not share a birthday with the first two is 363/365.

We want the second person not to share a birthday with you *and* the third person not to share a birthday with the first two people, so we use the multiplication rule:

$$P(\text{noshared birthday}) = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \approx 0.9918$$

and then subtract from 1 to get

P(shared birthday) = 1 – P(no shared birthday) = 1 – 0.9918 = 0.0082.

This is a pretty small number, so maybe it makes sense that the answer to our original problem will be small. Let's make our group a bit bigger.

Suppose five people are in a room. What is the probability that there is at least one shared birthday among these five people?

Solution:

Continuing the pattern of the previous example, the answer should be

$$P(\text{sharedbirthday}) = 1 - \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365} \cdot \frac{361}{365} \approx 0.0271$$

Note that we could rewrite this more compactly as

$$P(\text{sharedbirthday}) = 1 - \frac{_{365}P_5}{365^5} \approx 0.0271$$

which makes it a bit easier to type into a calculator or computer, and which suggests a nice formula as we continue to expand the population of our group.

Suppose 30 people are in a room. What is the probability that there is at least one shared birthday among these 30 people?

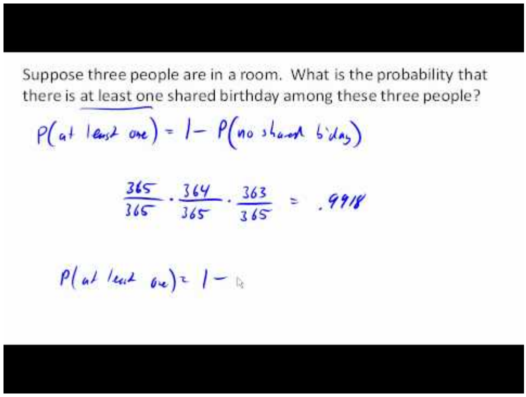
Solution:

Here we can calculate

$$P(\text{sharedbirthday}) = 1 - \frac{_{365}P_{30}}{365^{30}} \approx 0.706$$

which gives us the surprising result that when you are in a room with 30 people there is a 70% chance that there will be at least one shared birthday!

The birthday problem is examined in detail in the following.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

If you like to bet, and if you can convince 30 people to reveal their birthdays, you might be able to win some money by betting a friend that there will be at least two people with the same birthday in the room anytime you are in a room of 30 or more people. (Of course, you would need to make sure your friend hasn't studied probability!) You wouldn't be guaranteed to win, but you should win more than half the time.

This is one of many results in probability theory that is counterintuitive; that is, it goes against our gut instincts.

Try It

Suppose 10 people are in a room. What is the probability that there is at least one shared birthday among these 10 people?

Expected Value

Repeating Procedures Over Time

Expected value is perhaps the most useful probability concept we will discuss. It has many applications, from insurance policies to making financial decisions, and it's one thing that the casinos and government agencies that run gambling operations and lotteries hope most people never learn about.

example

In the casino game roulette, a wheel with 38 spaces (18 red, 18 black, and 2 green) is spun. In one possible bet, the player bets $1 on a single number. If that number is spun on the wheel, then they receive $36 (their original $1 + $35). Otherwise, they lose their $1. On average, how much money should a player expect to win or lose if they play this game repeatedly?

Solution:

Suppose you bet $1 on each of the 38 spaces on the wheel, for a total of $38 bet. When the winning number is spun, you are paid $36 on that number. While you won on that one number, overall you've lost $2. On a per-space basis, you have "won" -$2/$38 ≈ -$0.053. In other words, on average you lose 5.3 cents per space you bet on.

We call this average gain or loss the expected value of playing roulette. Notice that no one ever loses exactly 5.3 cents: most people (in fact, about 37 out of every 38) lose $1 and a very few people (about 1 person out of every 38) gain $35 (the $36 they win minus the $1 they spent to play the game).

There is another way to compute expected value without imagining what would happen if we play every possible space. There are 38 possible outcomes when the wheel spins, so the probability of winning is $\frac{1}{38}$. The complement, the probability of losing, is $\frac{37}{38}$.

Summarizing these along with the values, we get this table:

Outcome	Probability of outcome
$35	$\frac{1}{38}$
-$1	$\frac{37}{38}$

Notice that if we multiply each outcome by its corresponding probability we get $\$35 \cdot \frac{1}{38} = 0.9211$ and $-\$1 \cdot \frac{37}{38} = -0.9737$, and if we add these numbers we get

0.9211 + (-0.9737) ≈ -0.053, which is the expected value we computed above.

Expected Value

- **Expected Value** is the average gain or loss of an event if the procedure is repeated many times.

We can compute the expected value by multiplying each outcome by the probability of that outcome, then adding up the products.

Try It

You purchase a raffle ticket to help out a charity. The raffle ticket costs \$5. The charity is selling 2000 tickets. One of them will be drawn and the person holding the ticket will be given a prize worth \$4000. Compute the expected value for this raffle.

Example

In a certain state's lottery, 48 balls numbered 1 through 48 are placed in a machine and six of them are drawn at random. If the six numbers drawn match the numbers that a player had chosen, the player wins \$1,000,000. If they match 5 numbers, then win \$1,000. It costs \$1 to buy a ticket. Find the expected value.

Solution:

Earlier, we calculated the probability of matching all 6 numbers and the probability of matching 5 numbers:

$\frac{{}_6C_6}{{}_{48}C_6} = \frac{1}{12271512} \approx 0.0000000815$ for all 6 numbers,

$\frac{({}_6C_5)({}_{42}C_1)}{{}_{48}C_6} = \frac{252}{12271512} \approx 0.0000205$ for 5 numbers.

Our probabilities and outcome values are:

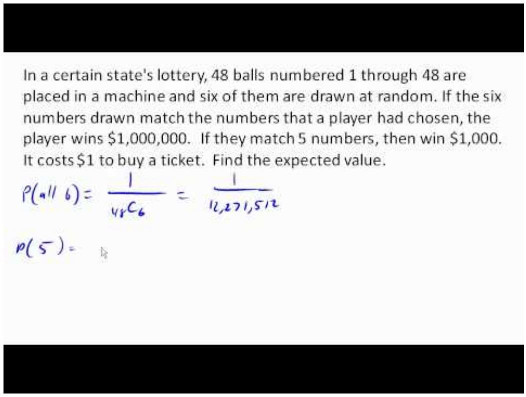
Outcome	Probability of outcome
\$999,999	$\frac{1}{12271512}$
\$999	$\frac{252}{12271512}$
-\$1	$1 - \frac{253}{12271512} = \frac{12271259}{12271512}$

The expected value, then is:

$$(\$999,999) \cdot \frac{1}{12271512} + (\$999) \cdot \frac{252}{12271512} + (-\$1) \cdot \frac{12271259}{12271512} \approx -\$0.898$$

On average, one can expect to lose about 90 cents on a lottery ticket. Of course, most players will lose \$1.

View more about the expected value examples in the following video.



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

Try It

Click here to try this problem.

In general, if the expected value of a game is negative, it is not a good idea to play the game, since on average you will lose money. It would be better to play a game with a positive expected value (good luck trying to find one!), although keep in mind that even if the *average* winnings are positive it could be the case that most people lose money and one very fortunate individual wins a great deal of money. If the expected value of a game is 0, we call it a **fair game**, since neither side has an advantage.

Try It

A friend offers to play a game, in which you roll 3 standard 6-sided dice. If all the dice roll different values, you give him $1. If any two dice match values, you get $2. What is the expected value of this game? Would you play?

Click here to try this problem.

Expected value also has applications outside of gambling. Expected value is very common in making insurance decisions.

Example

A 40-year-old man in the U.S. has a 0.242% risk of dying during the next year.[1] An insurance company charges $275 for a life-insurance policy that pays a $100,000 death benefit. What is the expected value for the person buying the insurance?

Solution:

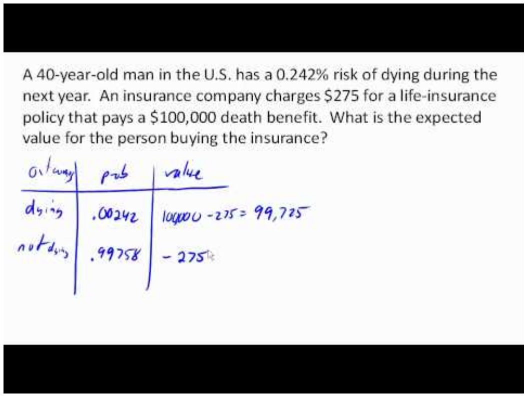
The probabilities and outcomes are

Outcome	Probability of outcome
$100,000 – $275 = $99,725	0.00242
-$275	1 – 0.00242 = 0.99758

The expected value is ($99,725)(0.00242) + (-$275)(0.99758) = -$33.

The insurance applications of expected value are detailed in the following video.

1. According to the estimator at http://www.numericalexample.com/index.php?view=article&id=91



A YouTube element has been excluded from this version of the text. You can view it online here: https://granite.pressbooks.pub/math502/?p=165

Not surprisingly, the expected value is negative; the insurance company can only afford to offer policies if they, on average, make money on each policy. They can afford to pay out the occasional benefit because they offer enough policies that those benefit payouts are balanced by the rest of the insured people.

For people buying the insurance, there is a negative expected value, but there is a security that comes from insurance that is worth that cost.

Attributions

PUTTING IT TOGETHER: PROBABILITY

The lottery jackpot has continued to climb as you completed this module. Now it is time to determine how likely you are to win.

Let's first assume that you not only need to pick six specific numbers from 1 – 49, but you need to pick them in the correct order. If this is the case, you know you need to use a permutation to figure out the size of the sample space.

$$P\left(n,r\right)=\frac{n!}{(n-r)!}$$

In this case, n is the possible numbers, which is 49, and r is the number of choices you make, which is 6.

$$P\left(49,6\right)=\frac{49!}{(49-6)!}$$

$$P\left(49,6\right)=\frac{49!}{43!}=10,068,347,520$$

This tells you that there is one way out of about 10 billion to win; your chances are not good at all.

Fortunately, most lottery winnings do not depend on order so you can use a combination instead.

$$C\left(n,r\right)=\frac{n!}{r!(n-r)!}$$
$$C\left(49,6\right)=\frac{49!}{6!(49-6)!}$$
$$C\left(49,6\right)=\frac{49!}{6!(43)!}$$
$$C\left(49,6\right)=\frac{49!}{6!(43)!}=13,983,816$$

Notice that the sample space has been greatly reduced from about 10 billion to about 14 million. So the likelihood of you winning is much greater than before, but still very slim.

What would happen to your chances of winning if you bought more than one ticket? Suppose you bought 100 tickets and chose a different combination of six numbers on each ticket. You could compare the number of tickets to sample space to determine your probability.

$$\frac{100}{14 \text{ million}} = 0.0000071 \quad = \quad 0.00071\%$$

That's much less than a 1% chance of winning. Still not very good. So suppose you gather up some cash and buy 1,000 tickets.

$$\frac{1{,}000}{14 \text{ million}} = 0.000071 \quad = \quad 0.0071\%$$

Now you are out $1000, assuming each ticket costs $1, and your chances are still less than a 1% chance.

Okay, maybe you are ready to go for broke. You and a group of friends gather your funds to purchase 1 million tickets.

$$\frac{1 \text{ million}}{14 \text{ million}} = 0.071 \quad = \quad 7.1\%$$

So even after purchasing 1 million tickets, which might cost $1 million, your probability of winning the big jackpot is only about 7%. To raise your probability to just 50%, you would have to purchase 7 million tickets. It's up to you do decide how lucky you feel. Maybe just buy one ticket and see what happens. Good luck!

Attributions

This chapter contains material taken from *Math in Society* (on [OpenTextBookStore]) by David Lippman, and is used

EXERCISES

Exercises

1. A ball is drawn randomly from a jar that contains 6 red balls, 2 white balls, and 5 yellow balls. Find the probability of the given event.
   a. A red ball is drawn
   b. A white ball is drawn
2. Suppose you write each letter of the alphabet on a different slip of paper and put the slips into a hat. What is the probability of drawing one slip of paper from the hat at random and getting:
   a. A consonant
   b. A vowel
3. A group of people were asked if they had run a red light in the last year. 150 responded "yes", and 185 responded "no". Find the probability that if a person is chosen at random, they have run a red light in the last year.
4. In a survey, 205 people indicated they prefer cats, 160 indicated they prefer dots, and 40 indicated they don't enjoy either pet. Find the probability that if a person is chosen at random, they prefer cats.
5. Compute the probability of tossing a six-sided die (with

sides numbered 1 through 6) and getting a 5.

6. Compute the probability of tossing a six-sided die and getting a 7.
7. Giving a test to a group of students, the grades and gender are summarized below. If one student was chosen at random, find the probability that the student was female.

	A	B	C	Total
Male	8	18	13	39
Female	10	4	12	26
Total	18	22	25	65

8. The table below shows the number of credit cards owned by a group of individuals. If one person was chosen at random, find the probability that the person had no credit cards.

	Zero	One	Two or more	Total
Male	9	5	19	33
Female	18	10	20	48
Total	27	15	39	81

9. Compute the probability of tossing a six-sided die and

getting an even number.

10. Compute the probability of tossing a six-sided die and getting a number less than 3.
11. If you pick one card at random from a standard deck of cards, what is the probability it will be a King?
12. If you pick one card at random from a standard deck of cards, what is the probability it will be a Diamond?
13. Compute the probability of rolling a 12-sided die and getting a number other than 8.
14. If you pick one card at random from a standard deck of cards, what is the probability it is not the Ace of Spades?
15. Referring to the grade table from question #7, what is the probability that a student chosen at random did NOT earn a C?
16. Referring to the credit card table from question #8, what is the probability that a person chosen at random has at least one credit card?
17. A six-sided die is rolled twice. What is the probability of showing a 6 on both rolls?
18. A fair coin is flipped twice. What is the probability of showing heads on both flips?
19. A die is rolled twice. What is the probability of showing a 5 on the first roll and an even number on the second roll?
20. Suppose that 21% of people own dogs. If you pick two people at random, what is the probability that they both own a dog?

21. Suppose a jar contains 17 red marbles and 32 blue marbles. If you reach in the jar and pull out 2 marbles at random, find the probability that both are red.
22. Suppose you write each letter of the alphabet on a different slip of paper and put the slips into a hat. If you pull out two slips at random, find the probability that both are vowels.
23. Bert and Ernie each have a well-shuffled standard deck of 52 cards. They each draw one card from their own deck. Compute the probability that:
    a. Bert and Ernie both draw an Ace.
    b. Bert draws an Ace but Ernie does not.
    c. neither Bert nor Ernie draws an Ace.
    d. Bert and Ernie both draw a heart.
    e. Bert gets a card that is not a Jack and Ernie draws a card that is not a heart.
24. Bert has a well-shuffled standard deck of 52 cards, from which he draws one card; Ernie has a 12-sided die, which he rolls at the same time Bert draws a card. Compute the probability that:
    a. Bert gets a Jack and Ernie rolls a five.
    b. Bert gets a heart and Ernie rolls a number less than six.
    c. Bert gets a face card (Jack, Queen or King) and Ernie rolls an even number.
    d. Bert gets a red card and Ernie rolls a fifteen.
    e. Bert gets a card that is not a Jack and Ernie rolls a

number that is not twelve.

25. Compute the probability of drawing a King from a deck of cards and then drawing a Queen.
26. Compute the probability of drawing two spades from a deck of cards.
27. A math class consists of 25 students, 14 female and 11 male. Two students are selected at random to participate in a probability experiment. Compute the probability that
    a. a male is selected, then a female.
    b. a female is selected, then a male.
    c. two males are selected.
    d. two females are selected.
    e. no males are selected.
28. A math class consists of 25 students, 14 female and 11 male. Three students are selected at random to participate in a probability experiment. Compute the probability that
    a. a male is selected, then two females.
    b. a female is selected, then two males.
    c. two females are selected, then one male.
    d. three males are selected.
    e. three females are selected.
29. Giving a test to a group of students, the grades and gender are summarized below. If one student was chosen at random, find the probability that the student was female and earned an A.

	A	B	C	Total
Male	8	18	13	39
Female	10	4	12	26
Total	18	22	25	65

30. The table below shows the number of credit cards owned by a group of individuals. If one person was chosen at random, find the probability that the person was male and had two or more credit cards.

	Zero	One	Two or more	Total
Male	9	5	19	33
Female	18	10	20	48
Total	27	15	39	81

31. A jar contains 6 red marbles numbered 1 to 6 and 8 blue marbles numbered 1 to 8. A marble is drawn at random from the jar. Find the probability the marble is red or odd-numbered.
32. A jar contains 4 red marbles numbered 1 to 4 and 10 blue marbles numbered 1 to 10. A marble is drawn at random from the jar. Find the probability the marble is blue or even-numbered.
33. Referring to the table from #29, find the probability that a student chosen at random is female or earned a B.
34. Referring to the table from #30, find the probability that

a person chosen at random is male or has no credit cards.

35. Compute the probability of drawing the King of hearts or a Queen from a deck of cards.
36. Compute the probability of drawing a King or a heart from a deck of cards.
37. A jar contains 5 red marbles numbered 1 to 5 and 8 blue marbles numbered 1 to 8. A marble is drawn at random from the jar. Find the probability the marble is
    a. Even-numbered given that the marble is red.
    b. Red given that the marble is even-numbered.
38. A jar contains 4 red marbles numbered 1 to 4 and 8 blue marbles numbered 1 to 8. A marble is drawn at random from the jar. Find the probability the marble is
    a. Odd-numbered given that the marble is blue.
    b. Blue given that the marble is odd-numbered.
39. Compute the probability of flipping a coin and getting heads, given that the previous flip was tails.
40. Find the probability of rolling a "1" on a fair die, given that the last 3 rolls were all ones.
41. Suppose a math class contains 25 students, 14 females (three of whom speak French) and 11 males (two of whom speak French). Compute the probability that a randomly selected student speaks French, given that the student is female.
42. Suppose a math class contains 25 students, 14 females (three of whom speak French) and 11 males (two of whom speak French). Compute the probability that a

randomly selected student is male, given that the student speaks French.

43. A certain virus infects one in every 400 people. A test used to detect the virus in a person is positive 90% of the time if the person has the virus and 10% of the time if the person does not have the virus. Let A be the event "the person is infected" and B be the event "the person tests positive".
    a. Find the probability that a person has the virus given that they have tested positive, i.e. find P(A | B).
    b. Find the probability that a person does not have the virus given that they test negative, i.e. find P(not A | not B).
44. A certain virus infects one in every 2000 people. A test used to detect the virus in a person is positive 96% of the time if the person has the virus and 4% of the time if the person does not have the virus. Let A be the event "the person is infected" and B be the event "the person tests positive".
    a. Find the probability that a person has the virus given that they have tested positive, i.e. find P(A | B).
    b. Find the probability that a person does not have the virus given that they test negative, i.e. find P(not A | not B).
45. A certain disease has an incidence rate of 0.3%. If the

false negative rate is 6% and the false positive rate is 4%, compute the probability that a person who tests positive actually has the disease.

46. A certain disease has an incidence rate of 0.1%. If the false negative rate is 8% and the false positive rate is 3%, compute the probability that a person who tests positive actually has the disease.
47. A certain group of symptom-free women between the ages of 40 and 50 are randomly selected to participate in mammography screening. The incidence rate of breast cancer among such women is 0.8%. The false negative rate for the mammogram is 10%. The false positive rate is 7%. If a the mammogram results for a particular woman are positive (indicating that she has breast cancer), what is the probability that she actually has breast cancer?
48. About 0.01% of men with no known risk behavior are infected with HIV. The false negative rate for the standard HIV test 0.01% and the false positive rate is also 0.01%. If a randomly selected man with no known risk behavior tests positive for HIV, what is the probability that he is actually infected with HIV?
49. A boy owns 2 pairs of pants, 3 shirts, 8 ties, and 2 jackets. How many different outfits can he wear to school if he must wear one of each item?
50. At a restaurant you can choose from 3 appetizers, 8 entrees, and 2 desserts. How many different three-course

meals can you have?

51. How many three-letter “words” can be made from 4 letters “FGHI” if
    a. repetition of letters is allowed
    b. repetition of letters is not allowed
52. How many four-letter “words” can be made from 6 letters “AEBWDP” if
    a. repetition of letters is allowed
    b. repetition of letters is not allowed
53. All of the license plates in a particular state feature three letters followed by three digits (e.g. ABC 123). How many different license plate numbers are available to the state’s Department of Motor Vehicles?
54. A computer password must be eight characters long. How many passwords are possible if only the 26 letters of the alphabet are allowed?
55. A pianist plans to play 4 pieces at a recital. In how many ways can she arrange these pieces in the program?
56. In how many ways can first, second, and third prizes be awarded in a contest with 210 contestants?
57. Seven Olympic sprinters are eligible to compete in the 4 x 100 m relay race for the USA Olympic team. How many four-person relay teams can be selected from among the seven athletes?
58. A computer user has downloaded 25 songs using an online file-sharing program and wants to create a CD-R with ten songs to use in his portable CD player. If the

order that the songs are placed on the CD-R is important to him, how many different CD-Rs could he make from the 25 songs available to him?

59. In western music, an octave is divided into 12 pitches. For the film *Close Encounters of the Third Kind*, director Steven Spielberg asked composer John Williams to write a five-note theme, which aliens would use to communicate with people on Earth. Disregarding rhythm and octave changes, how many five-note themes are possible if no note is repeated?
60. In the early twentieth century, proponents of the Second Viennese School of musical composition (including Arnold Schönberg, Anton Webern and Alban Berg) devised the twelve-tone technique, which utilized a tone row consisting of all 12 pitches from the chromatic scale in any order, but with not pitches repeated in the row. Disregarding rhythm and octave changes, how many tone rows are possible?
61. In how many ways can 4 pizza toppings be chosen from 12 available toppings?
62. At a baby shower 17 guests are in attendance and 5 of them are randomly selected to receive a door prize. If all 5 prizes are identical, in how many ways can the prizes be awarded?
63. In the 6/50 lottery game, a player picks six numbers from 1 to 50. How many different choices does the player have if order doesn't matter?

64. In a lottery daily game, a player picks three numbers from 0 to 9. How many different choices does the player have if order doesn't matter?
65. A jury pool consists of 27 people. How many different ways can 11 people be chosen to serve on a jury and one additional person be chosen to serve as the jury foreman?
66. The United States Senate Committee on Commerce, Science, and Transportation consists of 23 members, 12 Republicans and 11 Democrats. The Surface Transportation and Merchant Marine Subcommittee consists of 8 Republicans and 7 Democrats. How many ways can members of the Subcommittee be chosen from the Committee?
67. You own 16 CDs. You want to randomly arrange 5 of them in a CD rack. What is the probability that the rack ends up in alphabetical order?
68. A jury pool consists of 27 people, 14 men and 13 women. Compute the probability that a randomly selected jury of 12 people is all male.
69. In a lottery game, a player picks six numbers from 1 to 48. If 5 of the 6 numbers match those drawn, they player wins second prize. What is the probability of winning this prize?
70. In a lottery game, a player picks six numbers from 1 to 48. If 4 of the 6 numbers match those drawn, they player wins third prize. What is the probability of winning this prize?

71. Compute the probability that a 5-card poker hand is dealt to you that contains all hearts.
72. Compute the probability that a 5-card poker hand is dealt to you that contains four Aces.
73. A bag contains 3 gold marbles, 6 silver marbles, and 28 black marbles. Someone offers to play this game: You randomly select on marble from the bag. If it is gold, you win $3. If it is silver, you win $2. If it is black, you lose $1. What is your expected value if you play this game?
74. A friend devises a game that is played by rolling a single six-sided die once. If you roll a 6, he pays you $3; if you roll a 5, he pays you nothing; if you roll a number less than 5, you pay him $1. Compute the expected value for this game. Should you play this game?
75. In a lottery game, a player picks six numbers from 1 to 23. If the player matches all six numbers, they win 30,000 dollars. Otherwise, they lose $1. Find the expected value of this game.
76. A game is played by picking two cards from a deck. If they are the same value, then you win $5, otherwise you lose $1. What is the expected value of this game?
77. A company estimates that 0.7% of their products will fail after the original warranty period but within 2 years of the purchase, with a replacement cost of $350. If they offer a 2 year extended warranty for $48, what is the company's expected value of each warranty sold?
78. An insurance company estimates the probability of an

earthquake in the next year to be 0.0013. The average damage done by an earthquake it estimates to be \$60,000. If the company offers earthquake insurance for \$100, what is their expected value of the policy?

Exploration

Some of these questions were adapted from puzzles at mindyourdecisions.com.

79. A small college has been accused of gender bias in its admissions to graduate programs.
    a. Out of 500 men who applied, 255 were accepted. Out of 700 women who applied, 240 were accepted. Find the acceptance rate for each gender. Does this suggest bias?
    b. The college then looked at each of the two departments with graduate programs, and found the data below. Compute the acceptance rate within each department by gender. Does this suggest bias?

Department	Men		Women	
	Applied	Admitted	Applied	Admitted
Dept A	400	240	100	90
Dept B	100	15	600	150

c. Looking at our results from Parts *a* and *b*, what can you conclude? Is there gender bias in this college's admissions? If so, in which direction?

80. A bet on "black" in Roulette has a probability of 18/38 of winning. If you win, you double your money. You can bet anywhere from $1 to $100 on each spin.
    a. Suppose you have $10, and are going to play until you go broke or have $20. What is your best strategy for playing?
    b. Suppose you have $10, and are going to play until you go broke or have $30. What is your best strategy for playing?
81. Your friend proposes a game: You flip a coin. If it's heads, you win $1. If it's tails, you lose $1. However, you are worried the coin might not be fair coin. How could you change the game to make the game fair, without replacing the coin?
82. Fifty people are in a line. The first person in the line to have a birthday matching someone in front of them will win a prize. Of course, this means the first person in the line has no chance of winning. Which person has the highest likelihood of winning?
83. Three people put their names in a hat, then each draw a name, as part of a randomized gift exchange. What is the probability that no one draws their own name? What about with four people?

84. How many different "words" can be formed by using all the letters of each of the following words exactly once?
    a. "ALICE"
    b. "APPLE"
85. How many different "words" can be formed by using all the letters of each of the following words exactly once?
    a. "TRUMPS"
    b. "TEETER"
86. The *Monty Hall problem* is named for the host of the game show *Let's make a Deal*. In this game, there would be three doors, behind one of which there was a prize. The contestant was asked to choose one of the doors. Monty Hall would then open one of the other doors to show there was no prize there. The contestant was then asked if they wanted to stay with their original door, or switch to the other unopened door. Is it better to stay or switch, or does it matter?
87. Suppose you have two coins, where one is a fair coin, and the other coin comes up heads 70% of the time. What is the probability you have the fair coin given each of the following outcomes from a series of flips?
    a. 5 Heads and 0 Tails
    b. 8 Heads and 3 Tails
    c. 10 Heads and 10 Tails
    d. 3 Heads and 8 Tails
88. Suppose you have six coins, where five are fair coins, and one coin comes up heads 80% of the time. What is the

probability you have a fair coin given each of the following outcomes from a series of flips?

a. 5 Heads and 0 Tails
b. 8 Heads and 3 Tails
c. 10 Heads and 10 Tails
d. 3 Heads and 8 Tails

89. In this problem, we will explore probabilities from a series of events.

a. If you flip 20 coins, how many would you *expect* to come up "heads", on average? Would you expect *every* flip of 20 coins to come up with exactly that many heads?
b. If you were to flip 20 coins, what would you consider a "usual" result? An "unusual" result?
c. Flip 20 coins (or one coin 20 times) and record how many come up "heads". Repeat this experiment 9 more times. Collect the data from the entire class.
d. When flipping 20 coins, what is the theoretic probability of flipping 20 heads?
e. Based on the class's experimental data, what appears to be the probability of flipping 10 heads out of 20 coins?
f. The formula ${}_nC_x p^x(1-p)^{n-x}$ will compute the probability of an event with probability p occurring x times out of n, such as flipping x heads out of n coins where the probability of heads is p = ½. Use this to compute the theoretic probability of

flipping 10 heads out of 20 coins.

g. If you were to flip 20 coins, based on the class's experimental data, what range of values would you consider a "usual" result? What is the combined probability of these results? What would you consider an "unusual" result? What is the combined probability of these results?
h. We'll now consider a simplification of a case from the 1960s. In the area, about 26% of the jury eligible population was black. In the court case, there were 100 men on the juror panel, of which 8 were black. Does this provide evidence of racial bias in jury selection?

Attributions