Math in Society

Mathematics for liberal arts majors

Portland Community College

Edition 1.0

OpenOregon Educational Resources

Math in Society

Mathematics for liberal arts majors

Portland Community College Math Department

November 20, 2020

Project Lead: Cara Lee
Chapter Authors: Jess Brooks, Cara Lee, Sonya Redmond, Cindy Rochester-Gefre
Contributing Authors: Jess Brooks, Carlos Cantos, Teela Foxworth, Kandace Kling, Jeff Lacks, Cara Lee, Sonya Redmond, Cindy Rochester-Gefre, Thomas Songer, Eric Zupunski
Technology Editor: Emiliano Vega
Attribution: Many sections of this book are derivatives of Math in Society, by David Lippman et al., used under CC-BY-SA 3.0. See the Attributions page for details.

Edition: 1.0

Website: https://spot.pcc.edu/math/mathinsociety/

Notes

We dedicate this book to our students. May you have greater ease in paying for college and grow your proficiency and confidence in math.

Acknowledgements: We would like to thank Amy Hofer and grant support from Open Oregon Educational Resources and the PCC librarians and OER steering committee. Thanks also to Kaela Parks and Michael Cantino of Disability Services for their expertise on accessibility.

Antiracist and Culturally Responsive Elements: Our goal is for our students to see themselves in this book. We have added stories of mathematicians and economists with identities that have traditionally been underrepresented in textbooks. We use local places and culturally diverse names in the examples and exercises. We have examples with they/them/their pronouns and same-sex couples.

In the democracy chapter we have included the Native American genocide and slavery which are often left out of math texts. We have also emphasized the names of civil right leaders and African Americans killed by the police in our examples. We recognize there is more to do in this area and welcome contributions and feedback. Please email Cara Lee at cara.lee@pcc.edu or use the Google Form link below.

Philosophy: Our goal is for the content of this book to be relevant and accessible to our readers. We emphasize technology, conceptual understanding and communication over rote calculation. However, some manual calculation is important to understand what the technology is doing. We emphasize readily available spreadsheets and GeoGebra throughout the text.

Web and Print Versions: This book is available free online at http://spot.pcc.edu/math/mathinsociety.html. Chapters 1-5 are available at the PCC bookstore printed by the PCC Print Center.

Additional Topics: Your course may include one or more instructor choice topics which are available online only from the website above, or from David Lippman's original book at http://www.opentextbookstore.com/mathinsociety/index.html.

Accessibility: The html version of this book is intended to meet or exceed web accessibility standards. The alternate text for images is in progress. If you find anything that can improve the accessibility of this book, please email Cara Lee at cara.lee@pcc.edu or use the Google Form link below.

Online Homework: Online homework problems are available for free at https://www.myopenmath.com/.

Typos and Other Feedback: Please notify us of typos and other feedback using this Google Form[1].

[1] https://forms.gle/s14kNGd7anRmVgCD6

Attributions

Attributions:

- Cover Image: Portland, Oregon Skyline from the Ross Island Bridge, by Visitor7, used under CC BY-SA 3.0 Unported, cropped.
- Chapter 1 is a derivative of Math in Society: Logic, by David Lippman and Morgan Chase, and Math in Society: Sets by David Lippman, used under CC-BY-SA 3.0.
- Section 2.2, Section 2.3 and Section 2.4 are a derivative of Math in Society: Finance, by David Lippman, used under CC-BY-SA 3.0.
- Chapter 3 is a derivative of Math in Society: Describing Data and Statistics, by David Lippman, Jeff Eldridge and www.onlinestatbook.com, and www.onlinestatbook.com, by David M. Lane, et al, used under CC-BY-SA 3.0.
- Section 4.2 and Section 4.3 are a derivative of Math in Society: Probability, by David Lippman, used under CC-BY-SA 3.0.
- Section 5.1 is a derivative of Math in Society: Apportionment and Section 5.2 is a derivative of Math in Society: Voting Theory, by David Lippman, used under CC-BY-SA 3.0.
- Section 2.1, Section 2.5, Section 2.7 Credit Cards (online only), Section 4.1, Section 3.6 Normal Distribution (online only), Section 5.3, Section 5.4 and Section 5.6 Federal Deficit and National Debt (online only) are original to the Portland Community College Math Department.
- All chapter reviews are original to the Portland Community College Math Department.

- Images in the chapters that were not in the original OER or created by the PCC authors have attributions in footnotes.

Errata

These errors have been found since the last printing and fixed in the online version. The next printed books will have these corrections.

Table 0.0.1

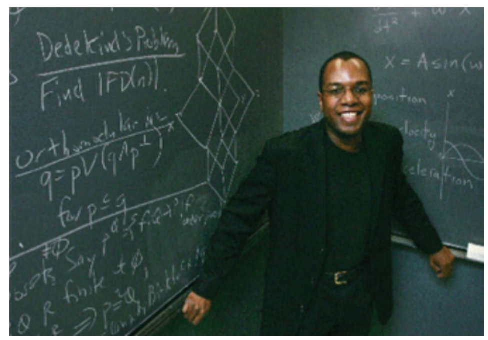
Section	*Error Description*
2.4	Example 2.4.5 was fixed to calculate 1% of the new balance. Example 2.4.6 was removed until it can be corrected.
5.1	Rodney King was not killed by the police but he was brutally beaten by four white police officers.

Contents

Chapter 1

Logic and Sets

Dr. Jonathan Farley and Partially Ordered Sets.



Dr. Jonathan Farley[1] grew up near Rochester, New York and got his mathematics degree from Harvard University. He then went to Oxford University in the United Kingdom. (Later he would return as a Fulbright Distinguished Scholar.) He was awarded Oxford's highest mathematics awards for graduate students, the Senior Mathematical Prize and the Johnson Prize, and earned his doctorate a year later (MIT, 2005). He had a two-year visit to the Mathematical Sciences Research Institute from 1995 to 1997. He is currently an Associate Professor at Morgan State University, a historically Black college.

Dr. Farley solved decades-old unsolved problems in the theory of ordered sets. Now he uses his expertise to help with counterterrorism. Dr. Farley applies lattice theory, a branch of mathematics that deals with ordered sets, to determine the probability that a terrorist cell has been disrupted after some of its members have been captured (Stafford, 2006). A group at Los Alamos National Laboratory used lattice theory for counterterrorism in the following way: A concept lattice can be drawn so that people who share many of the same characteristics are grouped together as one node, and links between nodes indicate that all the members of a subset with certain attributes must also have other attributes. Dr. Farley's work has been recognized both nationally and internationally and used by the Ministry of National Security in Jamaica (Hermoza, 2015).

Dr. Farley is also interested in making math more accessible and mentoring women in mathematics. He conceived of and co-organized a symposium on women and mathematics at Stanford University's Institute for Research on Women and Gender, called "Proof and Prejudice." He has co-founded Girls Equal Math, Equations of Peace, and an enterprise that creates math-themed fashion to encourage girls to go into math, Peren Linn Fashion (Brown, 2015).

Dr. Farley has also been a mathematics consultant for the hit TV shows *Numb3rs, Medium,* and *Elementary.* Dr. Farley's projects and applications of math continue to make it more interesting and relatable to a wider audience. You can read more on his website, Lattice Theory[2].

[1]Photo: ©Jonathan Farley. All rights reserved, used with permission.
[2]`http://www.latticetheory.net/`

1.1 The Language and Rules of Logic

Objectives: Section 1.1 The Language and Rules of Logic

Students will be able to:

- Identify propositions
- Compose and interpret the negation of a statement
- Use logical connectors (and/or) and conditional statements (if, then)
- Use truth tables to find truth values of basic and complex statements

1.1.1 Logic

Logic is the study of reasoning. Our goal in this chapter is to examine arguments to determine their validity and soundness. In this section we will look at propositions and logical connectors that are the building blocks of arguments. We will also use truth tables to help us examine complex statements.

1.1.2 Propositions

A ***propostion*** is a complete sentence that is either true or false. Opinions can be propositions, but questions or phrases cannot.

Example 1.1.1 Which of the following are propositions?

a I am reading a math book.

b Math is fun!

c Do you like turtles?

d My cat

Solution. The first and second items are propositions. The third one is a question and the fourth is a phrase, so they are not. We are not concerned right now about whether a statement is true or false. We will come back to that later when we examine full arguments. □

Arguments are made of one or more propositions (called ***premises***), along with a ***conclusion***. Propositions may be ***negated***, or combined with ***connectors*** like "and", and "or". Let's take a closer look at how these negations and logical connectors are used to create more complex statements.

1.1.3 Negation (not)

One way to change a proposition is to use its ***negation***, or opposite meaning. We often use the word "not" to negate a statement.

Example 1.1.2 Write the negation of the following propositions.

a I am reading a math book.

b Math is fun!

c The sky is not green.

d Cars have wheels

Solution.

a Negation: I am <u>not</u> reading a math book.

b Negation: Math is <u>not</u> fun!

c Negation: The sky is green (or <u>not</u> not green)

d Negation: Cars do <u>not</u> have wheels.

□

1.1.4 Multiple Negations

It is possible to use more than one negation in a statement. If you've ever said something like, "I can't not go," you are really saying you must go. It's a lot like multiplying two negative numbers which gives a positive result.

In the media and in ballot measures we often see multiple negations and it can be confusing to figure out what a statement means.

Example 1.1.3 Read the statement to determine the outcome of a yes vote.

"Vote for this measure to repeal the ban on plastic bags."

Solution. If you said that a yes vote would enable plastic bag usage, you are correct. The ban stopped plastic bag usage, so to repeal the ban would allow it again. This measure has a double negation and is also not very good for the environment. □

Example 1.1.4 Read the statement to determine the outcome on mandatory minimum sentencing.

"The bill that overturned the ban on mandatory minimum sentencing was vetoed."

Solution. In this case mandatory minimum sentencing would not be allowed. The ban would stop it, and the bill to overturn it was vetoed. This is an example of a triple negation. □

1.1.5 Logical Connectors (and, or)

When we use the word "and" between two propositions, it connects them to create a new statement that is also a proposition. For example, if you said "To finish this project, I need a screwdriver ***and*** a wrench," then you are expressing the need for both tools. For an "and" statement to be true, the connected propositions must both be true. If even one proposition is false (for instance, you didn't need a wrench) then the entire connected statement is false.

The word "or" between two propositions similarly connects the propositions to create a new statement. In this case, if you said "To finish this project, I need a screwdriver ***or*** a wrench," then you are expressing the need for one of the tools (but probably not both). For an "or" statement to be true, at least one of the propositions must be true (or both could be true).

1.1.6 Exclusive vs. Inclusive *or*

In English we often mean for *or* to be ***exclusive***: one or the other, but not both. In math, however, *or* is usually ***inclusive***: one or the other, or both. The thing we are including, or excluding is the "both" option.

Example 1.1.5 Determine whether each *or* statement is inclusive or exclusive.

a Would you like a chicken or vegan meal?

b We want to hire someone who speaks Spanish or Chinese

c Are you going to wear sandals or tennis shoes?

d Are you going to visit Thailand or Vietnam on your trip?

Solution. The first *or* statement is a choice of one or the other, but not both, so it is exclusive. The second statement is inclusive because they could find a candidate who speaks both languages. The third statement is exclusive because you can't wear both at the same time. The fourth statement is inclusive because you could visit both countries on your trip. □

1.1.7 Conditional Statements (if, then)

A ***conditional*** statement connects two propositions with *if, then.* An example of a conditional statement would be "*If* it is raining, *then* we'll go to the mall."

The statement "If it is raining," may be true or false for any given day. If the condition is true, then we will follow the course of action and go to the mall. If the condition is false, though, we haven't said anything about what we will or won't do.

1.1.8 Basic Truth Tables

In logic we can use a ***truth table*** to analyze a complex statement by summarizing all the possibilities and their ***truth values*** (true or false). To do this, we break the statement down to its smallest elements, the propositions. Then we can see the outcome of the complex statement for all possible combinations of true and false for the propositions.

For example, let's work with two propositions:

- R: You paid your rent this month.
- E: You paid your electric bill this month.

We will use these two propositions to demonstrate the truth tables for *not, and,* and *or.*

To set up a truth table, we list all the possible truth value combinations in a systematic way. The standard way of doing this is to make the first column half true, then half false, then cut the pattern in half with each succeeding column. For two propositions, the first two columns are shown to the right.

R	E	R *and* E
T	T	
T	F	
F	T	
F	F	

(a) Truth Table Setup for Two Propositions

Figure 1.1.6

The four possible combinations are

- Row 1: You have paid your rent and electric bill
- Row 2: You have paid your rent but not your electric bill
- Row 3: You have not paid your rent but you have paid your electric bill
- Row 4: You haven't paid either your rent or electric bill (yet).

Once we fill in the starting columns, we add additional columns for the more complex statements. We can add as many columns as needed. Below are the basic truth tables for *not, and,* and *or.*

Basic Truth Tables

Table 1.1.7 In the *not R* column, the truth value is the opposite of the value for *R*. For example, if *R* is true (you paid your rent) then *not R* (you did not pay your rent) is false.

Not

R	*not R*
T	F
F	T

Table 1.1.8 In the *R and E* column, you must have paid both your rent and electric bill. Otherwise *R and E* is false.

And

R	*E*	*R and E*
T	T	T
T	F	F
F	T	F
F	F	F

Table 1.1.9 In the *R or E* column, you must have paid either your rent or electric bill, or both (inclusive or). Otherwise *R or E* is false.

Or

R	*E*	*R or E*
T	T	T
T	F	T
F	T	T
F	F	F

1.1.9 Conditional Truth Tables

We talked about conditional statements (*if, then* statements), earlier. In logical arguments the first part (the "*if*" part) is usually a ***hypothesis*** and the second part (the "*then*" part) is a ***conclusion***.

To understand the truth table values for a conditional statement it is helpful to look at an example. Let's say a friend tells you, "If you post that photo to Facebook, you'll lose your job." Under what conditions can you say that your friend was wrong?

There are four possible outcomes:

1. You post the photo and lose your job
2. You post the photo and don't lose your job
3. You don't post the photo and lose your job
4. You don't post the photo and don't lose your job

The only case where you can say your friend was wrong is the second case, in which you post the photo but still keep your job.

Your friend didn't say anything about what would happen if you didn't post the photo, so you can't say the last two statements are wrong. Even if you didn't post the photo and lost your job anyway, your friend never said that you were guaranteed to keep your job if you didn't post it.

The four cases above correspond to the four rows of the truth table. For this truth table we will use P for "posting the photo," and L for "losing your job."

Table 1.1.10 Truth table for a conditional statement

P	L	If P, then L
T	T	T
T	F	F
F	T	T
F	F	T

If the hypothesis (the "if" part) is false, we cannot say that the statement is a lie, so the result of the third and fourth rows is true. Notice that we are using a double negation in this explanation.

We are using the words *and, or, not* and *if then* in this book, but if you look up other resources on truth tables you are likely to see these symbols.

Symbols used in other resources.

A and B is written $A \wedge B$

A or B is written $A \vee B$

not A is written A

If A, then B is written $A \to B$

1.1.10 Truth Tables for Complex Statements

Truth tables really become useful when we analyze more complex statements. In this case we will have several columns. It helps to work from the inside out and create a column in the table for each intermediate statement.

Example 1.1.11 Create a truth table for the statement A *or not* B

Solution.

When we create the truth table, we start with columns for the propositions, A and B. Then we add a column for *not* B because that is part of the final statement. Our last column is the final statement A *or not* B.

A	B	*not* B	A *or not* B
T	T		
T	F		
F	T		
F	F		

To complete the third column, *not* B, we take the opposite of the B column. Then to complete the fourth column, we only look at the A and the *not* B columns and compare them using *or*.

A	B	*not* B	A *or not* B
T	T	F	T
T	F	T	T
F	T	F	F
F	F	T	T

□

1.1.11 Truth Tables with Three Propositions

To create a truth table with three propositions we need eight rows for all the possible combinations. We will first determine the columns we need to get to our final statement. Then we will fill in the first three columns using the same methodology as before. Start with half true, half false, then cut the pattern in half each time.

Example 1.1.12 Create a truth table for the statement A *and not* (B *or* C)

Solution. First let's figure out the columns we will need. We have *A, B, C,* then we need the statement in the parentheses, *(B or C)*. Then we need the negation of that column, *not (B or C)*. Then we conclude with our final statement, *A and not (B or C)*.

Here is the initial table:

Table 1.1.13

A	*B*	*C*	*B or C*	*not (B or C)*	*A and not (B or C)*
T	T	T			
T	T	F			
T	F	T			
T	F	F			
F	T	T			
F	T	F			
F	F	T			
F	F	F			

Now we complete the columns one at a time. We use the *B* column and *C* column to complete *B or C,*. Then *not (B or C)* is the opposite of that column. For the final column we only need to look at the first and fifth columns, shaded in blue, with *and*. Here is the completed table.

Table 1.1.14

A	*B*	*C*	*B or C*	*not (B or C)*	*A and not (B or C)*
T	T	T	T	F	F
T	T	F	T	F	F
T	F	T	T	F	F
T	F	F	F	T	T
F	T	T	T	F	F
F	T	F	T	F	F
F	F	T	T	F	F
F	F	F	F	T	F

For this statement A must be true and neither B or C can be true, so it is only true in the fourth row. For an example of this statement, let's define these propositions in the context of professional baseball:

Let A = Anaheim wins, B = Baltimore wins, C = Cleveland wins.

Suppose that Anaheim will make the playoffs if: (1) Anaheim wins, and (2) neither Boston nor Cleveland wins. TFF is the only scenario in which Anaheim will make the playoffs. □

Example 1.1.15 Construct a truth table for the statement *if m and not p, then r.*

Solution. First, it may help to add parentheses to help you clarify the order. Our statement could also be written, *if (m and not p), then r.* To build this table, we will build the statement in parentheses and then repeat the *r* column after it. It's easier to read the conditional statement from left to right. Here are the columns for the table:

Table 1.1.16

m	p	r	*not p*	*m and not p*	r	*If (m and not p), then r*
T	T	T				
T	T	F				
T	F	T				
T	F	F				
F	T	T				
F	T	F				
F	F	T				
F	F	F				

For the fourth column, we take the opposite of p. Then we use the first and fourth columns to complete *m and not p*. With the r column repeated we can use columns five and six to complete our conditional statement. Here is the completed table:

Table 1.1.17

m	p	r	*not p*	*m and not p*	r	*If (m and not p), then r*
T	T	T	F	F	T	T
T	T	F	F	F	F	T
T	F	T	T	T	T	T
T	F	F	T	T	F	F
F	T	T	F	F	T	T
F	T	F	F	F	F	T
F	F	T	T	F	T	T
F	F	F	T	F	F	T

When m is true, p is false, and r is false—the fourth row of the table—then the hypothesis m and not p will be true, but the conclusion is false, resulting in an invalid conditional statement; every other case gives a true result.

If you want a real-life situation that could be modeled by if m and not p, then r, consider this:

Let $m =$ we order meatballs, $p =$ we order pasta, and $r =$ Ruba is happy.

The statement if *m and not p*, then r is, "if we order meatballs and don't order pasta, then Ruba is happy". If m is true (we order meatballs), p is false (we don't order pasta), and r is false (Ruba is not happy), then the statement is false, because we satisfied the premise, but Ruba did not satisfy the conclusion. ☐

In this section we have discussed propositions, logical connectors and truth tables. In the next section, we will look at set relationships before we analyze arguments.

1.1.12 Exercises

1. Which of the following are propositions?

a Pigs can fly.

b What?

c I don't know.

d I like tofu.

2. Which of the following are propositions?

a How far?

b Portland is not in Oregon.

c Portland Community College.

d It is raining.

3. Write the negation of each proposition.

a I ride my bike to campus.

b Portland is not in Oregon.

4. Write the negation of each proposition.

a You should see this movie.

b Lashonda is wearing blue.

5. Write a proposition that contains a double negative.

6. Write a proposition that contains a triple negative.

7. For each situation, decide whether the "or" is most likely exclusive or inclusive.

a An entrée at a restaurant includes soup or salad.

b You should bring an umbrella or a raincoat with you.

8. For each situation, decide whether the "or" is most likely exclusive or inclusive.

a We can keep driving on I-5 or get on I-405 at the next exit.

b You should save this document on your computer or a flash drive.

9. For each situation, decide whether the "or" is most likely exclusive or inclusive.

a. I will wear a sweater or a jacket.

b. My next vacation will be on the Oregon Coast or Mount Hood.

10. For each situation, decide whether the "or" is most likely exclusive or inclusive.

a. While in California I will go to the beach or Disneyland.

b. The insurance agent offers car or boat insurance.

11. Rewrite the statement in the conditional form *if p, then q.*

a. Whenever it is sunny, I go swimming.

b. I go see a movie on Fridays.

12. Rewrite the statement in the conditional form *if p, then q.*

a. I always carry an umbrella when it rains.

b. On the weekend I like to hang out with friends.

13. Translate each statement from symbolic notation into English sentences. Let A represent "Elvis is alive" and let K represent "Elvis is the King".

a *Not A*

b *A or K*

c *Not A and K*

d *If K, then not A*

14. Translate each statement from symbolic notation into English sentences. Let A represent "It rains in Oregon" and let B represent "I own an umbrella".

a *Not B*

b *A and not B*

c *If A, then B*

d *If not B, then A*

15. Translate each statement from English sentences into symbolic notation. Let A represent "I will protest" and let B represent "There is injustice."

a. There is injustice and I will protest.

b. If there is injustice, then I will protest.

c. I will protest if there is injustice.

d. If there is not injustice, then I will not protest.

16. Translate each statement from English sentences into symbolic notation. Let A represent "It's time to eat" and let B represent "I am hungry."

a. It's time to eat and I'm not hungry.

b. It's not time to eat.

c. If it's time to eat, then I'm hungry.

d. If I'm not hungry then it's not time to eat.

17. Determine if the entire statement is true or false.

a. An apple is a vegetable, or an apple is a fruit.

b. Portland is not a city in Oregon.

18. Determine if the entire statement is true or false.

a. Fish can walk and birds can swim.

b. If it is warm outside, then it is sunny.

Complete the truth table for each statement and write the meaning of each statement in the third column.

19. Let A be: I live in Oregon.

Let B be: I go to Portland Community College

A	B	A and B
T		
T		
F		
F		

20. Let A be: I am a psychology major

Let B be: I'm planning to transfer to Portland State

A	B	A or B
T		
T		
F		
F		

Complete the truth table for each statement.

21. *A and not B*

A	B	Not B	A and not B
T			
T			
F			
F			

22. *Not (not A or B)*

A	B	Not A	Not A or B	Not (not A or B)
T				
T				
F				
F				

23. *Not (A and B and C)*

A	B	C	A and B and C	Not (A and B and C)
T				
T				
T				
T				
F				
F				
F				
F				

24. *Not A or (not B and C)*

A	B	C	Not A	Not B	Not B and C	Not A or (Not B and C)
T						
T						
T						
T						
F						
F						
F						
F						

Create a complete truth table for each statement.

25. *Not(A and B) or C*

26. *(A or B) and (A or C)*

27. *If (A and B), then C*

28. *If (A or B), then not C*

29. *If (A and C), then not A*

30. *If (B or C), then (A and B)*

1.2 Sets and Venn Diagrams

Objectives: Section 1.2 Sets and Venn Diagrams

Students will be able to:

- Use set notation and understand the null set
- Determine the universal set for a given context
- Use Venn diagrams and set notation to illustrate the intersection, union and complements of sets
- Illustrate disjoint sets, subsets and overlapping sets with diagrams
- Use Venn diagrams and problem-solving strategies to solve logic problems

1.2.1 Sets

It is natural for us to classify items into groups, or ***sets***, and consider how they interact with each other. In this section, we will use sets and Venn diagrams to visualize relationships between groups and represent survey data.

A ***set*** is a collection of items or things. Each item in a set is called a ***member*** or an ***element***.

Example 1.2.1

a. The numbers 2 and 42 are elements of the set of all even numbers.

b. MTH 105 is a member of the set of all courses you are taking.

□

A set consisting entirely of elements of another set is called a ***subset***. For instance, the set of numbers 2, 6, and 10 is a subset of the set of all even numbers.

Some sets, like the set of even numbers, can be defined by simply describing their contents. We can also define a set by listing its elements using ***set notation***.

1.2.2 Set Notation

Set notation is used to define the contents of a set. Sets are usually named using a capital letter, and its elements are listed *once* inside a set of curly brackets.

For example, to write the set of primary colors using set notation, we could name the set C for colors, and list the names of the primary colors in brackets: $C = \{red, yellow, blue\}$. In this case, the set C is a subset of all colors. If we wanted to write the list of our favorite foods using set notation, we could write $F = \{cheese, raspberries, wine\}$. And yes, wine is definitely an element of some food group!

Example 1.2.2 Julia, Keenan, Jae and Colin took a test. They got the following scores: 70, 95, 85 and 70. Let P be the set of test takers and S be the set of test scores. List the elements of each set using set notation.

Solution. In this example, the set of people taking the test is $P = \{\text{Julia, Keenan, Jae, Colin}\}$, and the set of test scores is $S = \{70, 85, 95\}$. Notice in this example that even though two people scored a 70 on the test, the score of 70 is only listed once. □

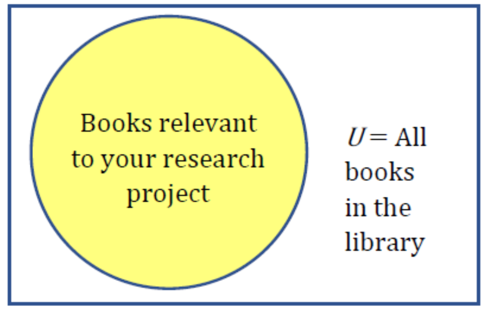
It is important to note that when we write the elements of a set in set notation, there is no order implied. For example, the set $\{1, 2, 3\}$ is equivalent to the set $\{3, 1, 2\}$. It is conventional, however, to list the elements in order if there is one.

1.2.3 Universal Set

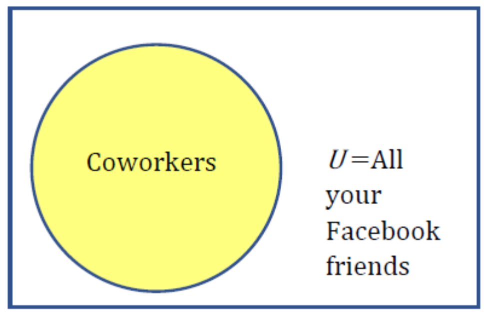
The ***universal set*** is the set containing every possible element of the described context. Every set is therefore a subset of the universal set. The universal set is often illustrated by a rectangle labeled with a capital letter U. Subsets of the universal set are usually illustrated with circles for simplicity, but other shapes can be used.

Example 1.2.3

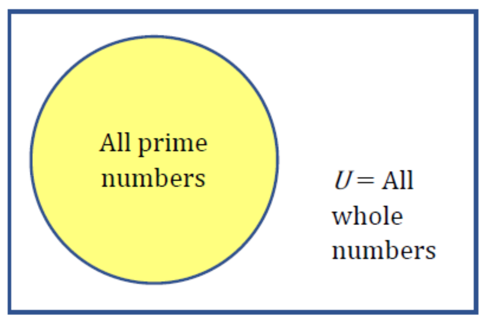
a. If you are searching for books for a research project, the universal set might be all the books in the library, and the books in the library that are relevant to your research project would be a subset of the universal set.



b. If you are wanting to create a group of your Facebook friends that are coworkers, the universal set would be all your Facebook friends and the group of coworkers would be a subset of the universal set.



c. If you are working with sets of numbers, the universal set might be all whole numbers, and all prime numbers would be a subset of the universal set.



□

1.2.4 The Null Set

It is possible to have a set with nothing in it. This set called the ***null set*** or ***empty set***. It's like going to the grocery store to buy your favorite foods and realizing you left your wallet at home. You walk away with an empty bag. The set of items that you bought at the grocery store would written in set notation as $G = \{ \}$, or $G = \emptyset$.

1.2.5 Intersection, Union, and Complement (And, Or, Not)

Suppose you and your roommate decide to have a house party, and you each invite your circle, or set, of friends. When you combine your two sets of friends, you discover that you have some friends in common.

The set of friends that you have in common is called the ***intersection***. The ***intersection*** of two sets contains only the elements that are in both sets. To be in the intersection of set A and B, an element needs to be in both set A ***and*** set B.

The set of all friends that you and your roommate have invited is called the ***union***. The ***union*** of two sets contains all the elements contained in either set (or both). To be in the union of set A and set B, an element must to be contained in just set A, just set B, ***or*** in the intersection of sets A and B. Notice that in this case that the "or" is inclusive.

What about the people who were *not* invited to the party and showed up anyway? They are not elements of your set of invited friends. Nor are they an element of your roommate's set of invited friends. These uninvited party crashers are the ***complement*** to your set of invited friends. The complement of a set A contains everything that is *not* in the set A. To be in the complement of set A, an element ***cannot*** be in set A, but it will be an element of the universal set.

Example 1.2.4 Consider the sets: A = {red, green, blue}, B = {red, yellow, orange}, and C = {red, orange, yellow, green, blue, purple}

a. Determine the set A intersect B, and write it in set notation.

b. Determine the set A union B, and write it in set notation.

c. Determine the intersection of A complement and C and write it in set notation.

Solution.

a. The intersection contains the elements in both sets: A intersect B = {red}

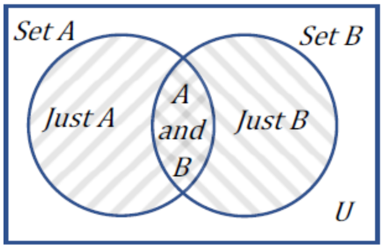
b. The union contains all the elements in either set: A union B = {red, green, blue, yellow, orange}. Notice we only list red once.

c. Here we are looking for all the elements that are *not* in set A and are in set C: A complement intersect C = {orange, yellow, purple}

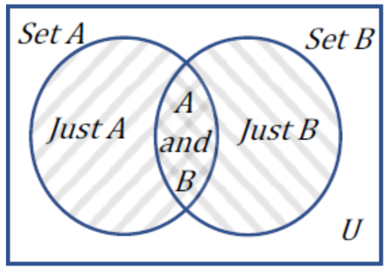
□

1.2.6 Venn Diagrams

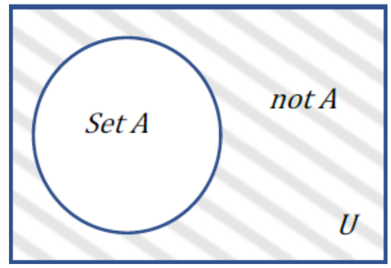
Venn diagrams are used to illustrate the relationships between two or more sets. To create a Venn diagram, start by drawing a rectangle to represent the universal set. Next draw and label overlapping circles to represent each of your sets. Most often there will be two or three sets illustrated in a Venn diagram. Finally, if you are given elements, fill in each region with its corresponding elements. Venn diagrams are also a great way to illustrate intersections, unions and complements of sets as shown below.



(a) The ***intersection*** is where the shading of the two sets overlaps in the center. It contains the elements of A ***and*** B.



(b) The ***union*** includes all elements of A ***or*** B or both. It contains all three of the shaded regions.



(c) The ***complement*** of set A includes all the elements ***not*** in A. It is the shaded region outside the set of A, but within the universal set.

Figure 1.2.5

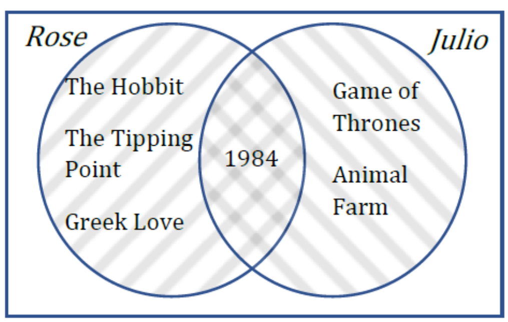
Here is an example of how to draw a Venn Diagram.

Example 1.2.6 Let J be the set of books Julio read this summer and let R be the set of books Rose read this summer. Draw a Venn diagram to show the sets of books they read if Julio read Game of Thrones, Animal Farm and 1984, and Rose read The Hobbit, 1984, The Tipping Point, and Greek Love.

To create a Venn diagram showing the relationship between the set of books Julio read and the set of books Rose read, first draw a rectangle to illustrate the universal set of all books.

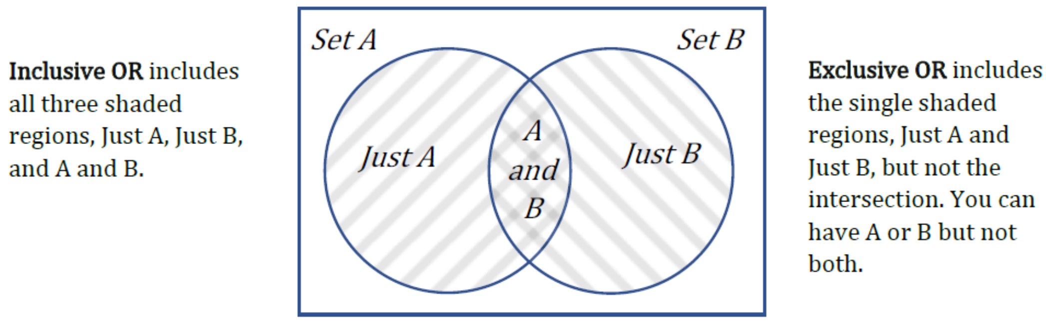
Next draw two overlapping circles, one for the set of books Julio read and one for the set of books Rose read. Since both Rose and Julio read 1984, we place it in the overlapping region (the intersection).

All the books that Rose read will lie in her circle, in one of the two regions that make up her set. Likewise for the books Julio read. Since we have already filled in the overlapping region, we put the books that only Rose read in her circle's "cresent moon" section, and we put the books that only Julio read in his circle's "cresent moon" section. The resulting diagram is shown below.



□

Example 1.2.7 In the last section we discussed the difference between ***inclusive "or"*** and ***exclusive "or."*** In common language, "or" is usually exclusive, meaning the set A or B includes just A or just B but not both. In logic, however, "or" is inclusive, so the set A or B includes just A, just B, or both. The difference between the inclusive and exclusive "or" can be illustrated in a Venn, as shown below.



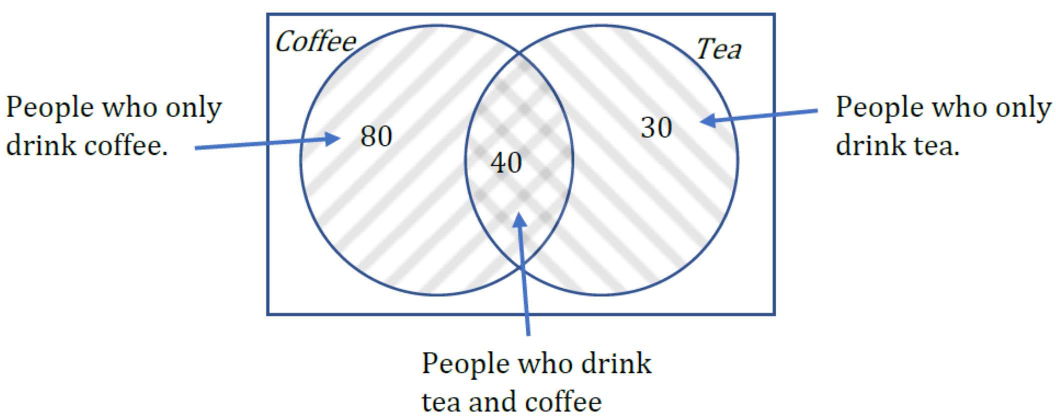
□

1.2.7 Illustrating Data

We can also use Venn diagrams to illustrate quantities, data, or frequencies.

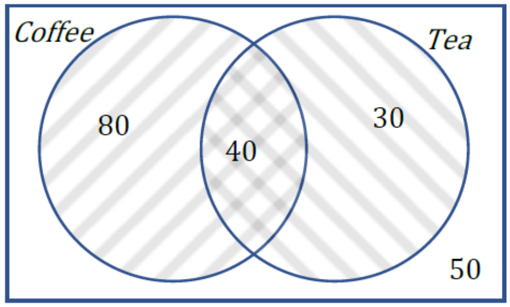
Example 1.2.8 A survey asks 200 people, "What beverage(s) do you drink in the morning?" and offers three choices: tea only, coffee only, and both coffee and tea. Thirty report drinking only tea in the morning, 80 report drinking only coffee in the morning, and 40 report drinking both. How many people drink tea in the morning? How many people drink neither tea nor coffee?

Solution. To answer this question, let's first create a Venn diagram representing the survey results. Placing the given values, we have the following:



The universal set should include all 200 people surveyed, but we only have 150 placed so far. The difference between what we have placed so far, and the 200 total is the number of people who drink neither coffee nor tea. These $200 - 150 = 50$ people are placed outside of the circles but within the rectangle since they are still included in the universal set.

The number of people who drink tea in the morning includes everyone in the tea circle. This includes those who only drink tea and those who drink both tea and coffee. Thus, the number of people who drink tea is $40 + 30 = 70$.



□

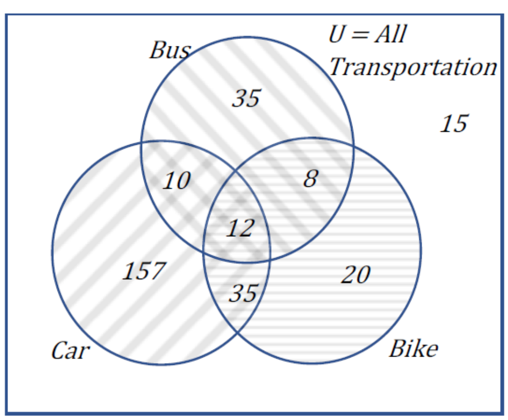
Here is an example of a Venn diagram with three sets.

Example 1.2.9 In a survey, adults were asked how they travel to work. Below is the recorded data on how many people took the bus, biked, and/or drove to work. Draw and label a Venn diagram using the information in the table.

Travel Options	***Frequency***
Just Car	157
Just Bike	20
Just Bus	35
Car and Bike only	35
Car and Bus only	10
Bus and Bike only	8
Car, Bus and Bike	12
Neither Car, Bus nor Bike	15
Total	292

Solution. To fill in the Venn diagram, we will place the 157 people who only drive a car in the car set where it does not overlap with any other modes of transportation. We can fill in the numbers 20 and 35 in a similar way.

Then we have the overlap of two modes of transportation only. There are 35 people who use their car and bike only, so they go in the overlap of those two sets, but they do not take the bus, so they are outside of the bus set. Similarly, we can enter the 10 and 8. There are 12 people who use all three modes, so they are in the intersection of all three sets. There are 15 people who do not use any of the three modes, so they are placed outside the circles but inside the universal set of all modes of transportation. Here is the completed Venn diagram.



□

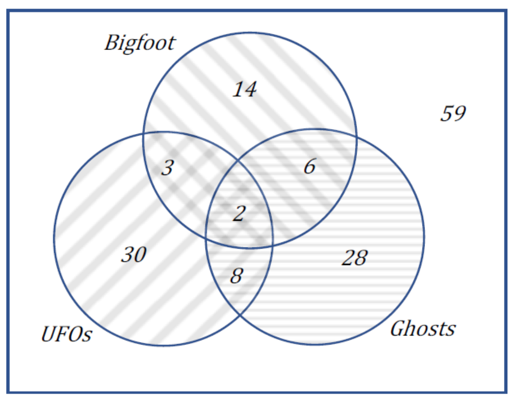
Example 1.2.10 One hundred fifty people were surveyed and asked if they believed in UFOs, ghosts, and Bigfoot. The following results were recorded.

- 43 believed in UFOs
- 44 believed in ghosts
- 25 believed in Bigfoot
- 10 believed in UFOs and ghosts
- 8 believed in ghosts and Bigfoot
- 5 believed in UFOs and Bigfoot
- 2 believed in all three

Draw and label a Venn diagram to determine how many people believed in at least two of these things.

Solution. Starting with the intersection of all three circles, we work our way out. The number in the center is 2, since two people believe in UFO's, ghosts and Bigfoot. Since 10 people believe in UFOs and Ghosts, and that includes the 2 that believe in all three, that leaves 8 that believe in only UFOs and Ghosts.

We work our way out, filling in all the regions. Once we have, we can add up all those regions, getting 91 people in the union of all three sets. This leaves $150 - 91 = 59$ who believe in none.



Then to answer the question of how many people believed in ***at least two*** (two or more), we add up the numbers in the intersections, $8 + 2 + 3 + 6 = 19$ people. □

1.2.8 Qualified Propositions

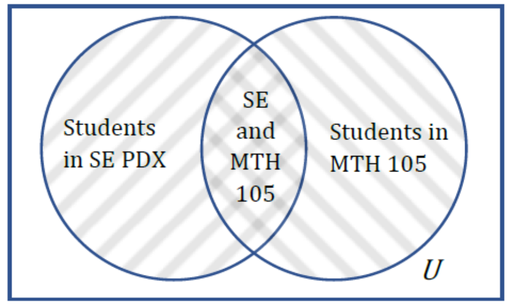
A ***qualified proposition*** is a statement that asserts a relationship between two sets. The three relationships we will be looking at in this section are "some" (some elements are shared between the two sets), "none" (none of the elements are shared between the two sets), and "all" (all elements of one set are contained in the other set). These relationships are especially important in evaluating arguments.

1.2.9 Overlapping Sets

Sets overlap if they have members in common. The Venn diagram examples we have looked at in this section are ***overlapping*** sets.

Example 1.2.11

The set of students living in SE Portland and the set of students taking MTH 105.
Qualified Proposition: "***Some*** students who live in SE Portland take MTH 105."



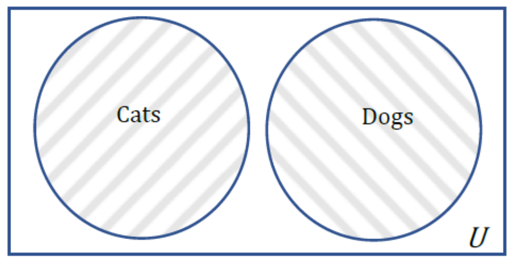
□

1.2.10 Disjoint Sets

Sets are ***disjoint*** if they have no members in common.

Example 1.2.12

The set of Cats and the set of Dogs.
Qualified Proposition: "***No*** cats are Dogs."



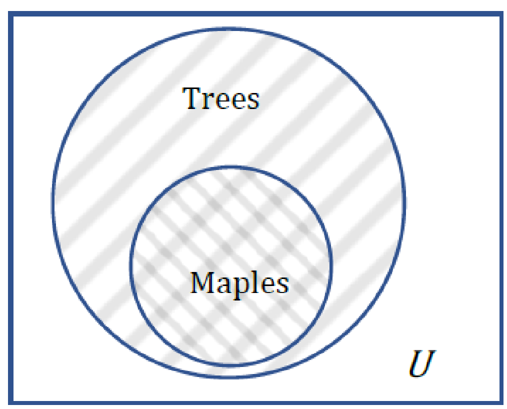
□

1.2.11 Subsets

If a set is completely contained in another set, it is called a ***subset***.

Example 1.2.13

The set of all Trees and the set of Maples Trees.
Qualified Proposition: "***All*** Maples are Trees."



□

1.2.12 Exercises

1. List the elements of the set "The letters of the word Mississippi."

2. List the elements of the set "Months of the year."

3. Write a verbal description of the set {3, 6, 9}.

4. Write a verbal description of the set {a, i, e, o, u}.

5. Is {1, 3, 5} a subset of the set of odd numbers?

6. Is {A, B, C} a subset of the set of letters of the alphabet?

Create a Venn diagram to illustrate each of the following:

7. A survey was given asking whether people watch movies at home from Netflix, Redbox, or Disney+. Use the results to determine how many people use Redbox.

- 70 only use Netflix, 30 only use Redbox
- 5 only use Disney+, 6 use only Disney+ and Redbox
- 14 use only Netflix and Redbox, 20 use only Disney+ and Netflix
- 7 use all three, 25 use none of these

8. A survey asked buyers whether color, size, or brand influenced their choice of cell phone. The results are below. How many people were influenced by brand?

- 5 said only color, 8 said only size
- 16 said only brand, 20 said only color and size
- 42 said only color and brand, 53 said only size and brand
- 102 said all three, 20 said none of these

9. Use the given information to complete a Venn diagram, then determine: a) how many students have seen exactly one of these movies, and b) how many have seen only Star Wars Episode IX.

- 25 have seen *Inception (I)*, 45 have seen *Star Wars Episode IX (SW)*
- 19 have seen *Vanilla Sky (VS)*, 18 have seen *I* and *SW*
- 17 have seen *VS* and *SW*, 11 have seen *I* and *VS*
- 9 have seen all three
- 2 have seen none of these

10. A survey asked people what alternative transportation modes they use. Use the data to complete a Venn diagram, then determine: a) what percentage of people only ride the bus, and b) how many people don't use any alternate transportation.

- 40 use the bus, 25 ride a bicycle
- 33 walk, 7 use the bus and ride a bicycle
- 15 ride a bicycle and walk, 20 use the bus and walk
- 5 use all three

Given the qualified propositions:

a. Determine the two sets being described.

b. Determine if the sets described are Subsets, Overlapping Sets or Disjoint sets.

c. illustrate the situation using sets.

11. All Terriers are dogs.

12. Some Mammals Swim. (The second set is not clearly defined but is implied)

13. No pigs can fly.

14. All children are young.

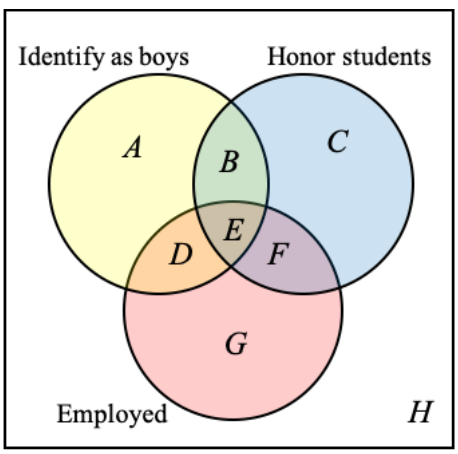
15. Some friends remember your birthday.

16. No lies are truths.

17. Suppose someone wants to conduct a study to learn how teenage employment rates differ among gender identities and honor roll status. For each teenager in the study, they will need to record the answers to these three questions:

- How does the teenager identify their gender?
- Is the teenager an honor student or not?
- Is the teenager employed or not?

Explain what each of the regions in the following Venn Diagram represent.



- Region A:_________
- Region B:_________
- Region C:_________
- Region D:_________
- Region E:_________
- Region F:_________
- Region G:_________
- Region H:_________

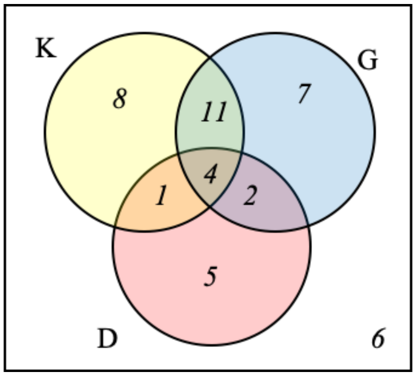
18. Students were surveyed to see if they used smart phones, tablets, or both.

- 85 students said they used smart phones
- 50 students said they used tablets
- 10 students said they use both
- 5 students said they used neither

Draw a Venn Diagram to help you answer the following. Show your calculations.

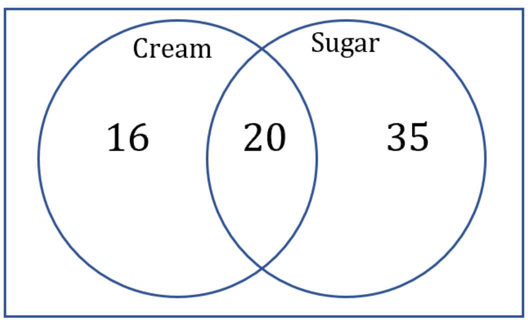
How many students were surveyed?

19. The Venn Diagram below shows the numbers of students in grade 5 at an elementary school who have enrolled for keyboard (K), guitar (G), and drum classes. Fill in the blanks



- Students enrolled for keyboard class:_________
- Students enrolled for keyboard class only:_________
- Students didn't enroll at all:_________
- Students took all three classes:_________
- Students enrolled for guitar and drum:_________
- Students enrolled for guitar and drum only:_________

20. A poll asked 100 coffee drinkers whether they like cream or sugar in their coffee. The information was organized in the following Venn Diagram.



a How many coffee drinkers like cream?

b How many coffee drinkers like sugar?

c How many coffee drinkers like sugar but not cream?

d How many coffee drinkers like cream but not sugar?

e How many coffee drinkers like cream and sugar?

f How many coffee drinkers like cream or sugar?

g How many coffee drinkers like neither cream nor sugar?

21. 110 dogs were asked "Why do you like to eat garbage?"

- 89 said "It tastes great!"
- 87 said "It's more filling!"
- 68 said "It tastes great!" and "It's more filling!"

a. Draw a Venn diagram to represent this information

b. How many said "It's more filling!" but didn't say "It tastes great!"?

c. How many said neither of those things?

d. How many said "It's more filling!" or said "It tastes great"

1.3 Describing and Critiquing Arguments

Objectives: Section 1.3 Describing and Critiquing Arguments.

Students will be able to:

- Understand the structure of logical arguments by identifying the premise(s) and conclusion
- Distinguish between inductive and deductive arguments
- Make a set diagram to evaluate deductive arguments
- Determine whether a deductive argument is valid and/or sound

1.3.1 Logical Arguments

A ***logical argument*** is a claim that a set of ***premises*** support a ***conclusion***. It is possible for a logical argument to have one or many premises, but there must be one conclusion. In this section we will look at types of arguments and how to determine the strength, validity and/or soundness of each type. There are two types of arguments we will explore in this section: ***inductive*** and ***deductive*** arguments.

1.3.2 Inductive and Deductive Arguments

To better understand the difference between inductive and deductive arguments, let's start by looking at a couple of examples.

Example 1.3.1 Consider the following argument:

When I went to the store last week, I forgot my wallet, and I forgot it again when I went back today. I always forget my wallet when I go to the store.

Before we analyze an argument, it is helpful to precisely state its premises and its conclusion. Most arguments you encounter in the real world won't be stated in a precise "premise, premise, conclusion" form. Sometimes the conclusion will be stated before the premises, or the premises will be hidden within a bunch of rhetoric.

To begin our analysis of this first argument, let's first rewrite it in a more precise "premise, premise, conclusion" form.

Premise: I forgot my wallet when I went to the store last week.

Premise: I forgot my wallet when I went to the store today.

Conclusion: I always forget my wallet when I go to the store.

Notice that both premises make a claim about a ***specific*** instance – the specific instance last week when I forgot my wallet, and the specific instance today when I forgot my wallet. The conclusion, on the other hand, states what we can expect to happen more generally. □

Now let's consider a different argument:

Example 1.3.2 Henry must know CPR because he is a nurse and all nurses know CPR.

Just as we did for the last example, let's rewrite the argument in its "premise, premise, conclusion" form:

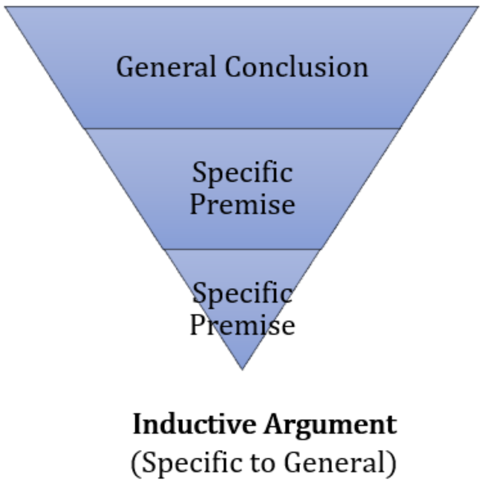
Premise: All nurses know CPR.

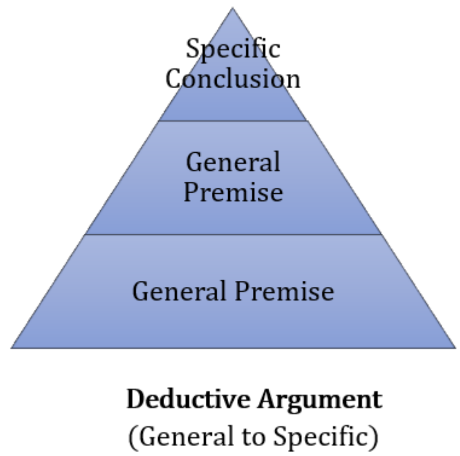
Premise: Henry is a nurse.

Conclusion: Henry knows CPR.

Unlike the first argument where the premises were specific and the conclusion was general, this argument's first premise is a general statement and the conclusion is specific. We can determine whether an argument is inductive or deductive by looking at which part of the argument is general and which is specific. In the first example, the premises were specific and the conclusion was more general. This is an example of an ***inductive*** argument. In the second example, it was the premises that were more general and the conclusion that was specific. This is an example of a ***deductive*** argument. □

In general, an ***inductive argument*** uses a collection of *specific* examples (i.e. data) as its premises and uses them to propose a *general* conclusion, while a ***deductive argument*** uses a collection of *general statements* (i.e. definitions) as its premises and uses them to propose a *specific conclusion.* You can see the difference in the pyramids below. We start with the premises at the bottom and build up to the conclusion.





Example 1.3.3 Rewrite the following arguments in a precise "premise, premise, conclusion" form, and determine if the argument is inductive or deductive.

a. A number is prime if it is only divisible by itself and one. Since the number 13 is only divisible by itself and one, 13 must be prime.

b. Juan's dog Goober is having puppies. All three of Goober's previous litters have had 5 puppies so Goober is bound to have 5 puppies in this litter as well.

Solution.

a. Premise: If a number is only divisible by itself and one, the number is prime.

   Premise: The number 13 is only divisible by itself and one.

   Conclusion: 13 is prime.

   Since the premises are general definitions and properties of numbers and the conclusion is a specific statement about the number 13, the argument is ***deductive***.

b. Premise: Goober is having puppies.

   Premise: Goober's last three litters had 5 puppies.

   Conclusion: Goober's current litter will have 5 puppies.

   This is an example of an ***inductive argument*** since it uses specific experiences/instances as its premises, and its conclusion is a general expectation based on those specific experiences.

□

1.3.3 Evaluating Arguments

Inductive arguments cannot be proven. The best we can do is evaluate the ***strength*** of the argument based on the evidence it provides.

A strong inductive argument is one that is well supported by its premises, while a weak inductive argument is one whose premises do a poor job of supporting the conclusion. The strength of an inductive argument is subjective, because where one person sees a strong argument, another may see a weak argument. Additionally, the strength and truth of an argument are not necessarily related; it is possible to have a weak argument that is true, and a strong argument that is false.

Example 1.3.4 Determine the strength of the inductive argument.

James Franco, Jodie Foster, Jennifer Lawrence, and Jack Nicholson have all won Academy Awards for acting. Actors whose names start with J are bound to win an Academy Award.

Solution. The inductive argument provides a number of specific cases as evidence for the conclusion. However, we would not be surprised if a J-named actor *did not* win an Academy Award, so the argument is weak. □

Deductive arguments, on the other hand, can be proven and their validity and soundness can be evaluated. The ***validity of the argument*** is based on whether the conclusion follows logically from the premises, while the ***soundness of the argument*** is based on whether or not the premises are true. An argument cannot be sound if it is not valid, even if the premises seem reasonable.

1.3.4 Evaluating Deductive Arguments Using Sets

One way to determine whether a deductive argument is valid is to illustrate the premises of the argument using sets and see if the conclusion logically follows if we assume the premises to be true.

Example 1.3.5 Use a set diagram to determine whether the argument is valid. If the argument is valid, determine if it is also sound.

a. "All cats are mammals and a tiger is a cat, so a tiger is a mammal."

b. "All water bottles are plastic. This is a water bottle, so it must be plastic."

c. "All firefighters know CPR. Jill knows CPR, so Jill must be a firefighter."

d. "None of my friends like dancing. Kai doesn't like dancing. Therefore, Kai is my friend."

e. "Some young adults make minimum wage and Tara is a young adult. Therefore, Tara makes minimum wage."

Solution.

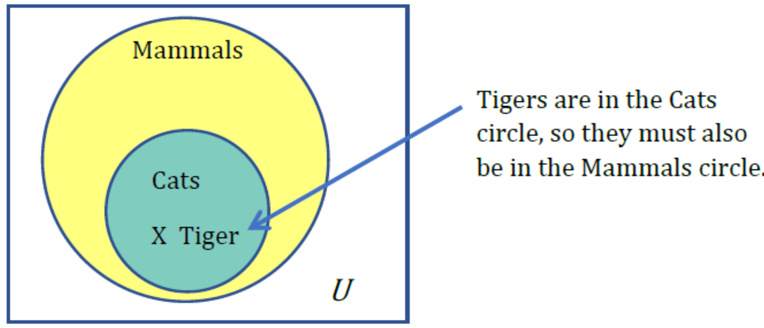
a. First let's write the argument in its "premise, premise, conclusion" form. For the problems we will be looking at, you will want to write the first premise as a ***qualified proposition*** (some, none, all) since this will form the basic structure of our diagram.

   Premise: All cats are mammals.

   Premise: A tiger is a cat.

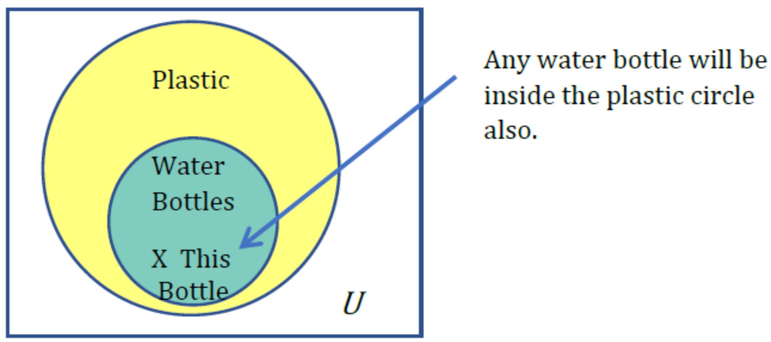
   Conclusion: A tiger is a mammal.

   From the first premise we know that all cats lie inside the set of mammals (cats are a subset of mammals). From the second premise, we know that tigers lie inside the set of cats (marked with an X), and therefore also lie within the set of mammals.



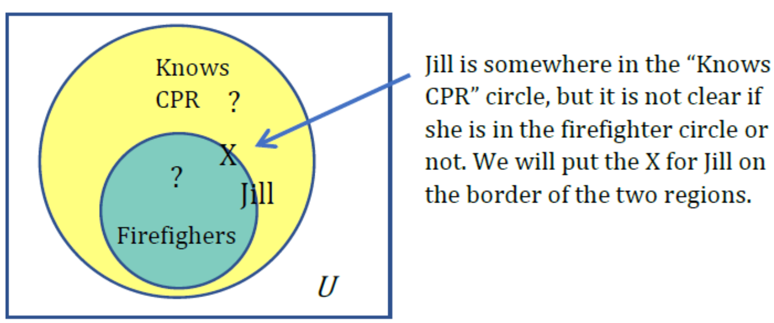
This argument is valid because we were able to show that the conclusion follows logically from the premises. The argument is also sound since the premises "all cats are mammals" and "a tiger is a cat" are true.

b. From the first premise we know that all water bottles lie inside the set of plastic items (water bottles are a subset of plastic). From the second premise, we know that this particular water bottle must lie within the plastic items set.

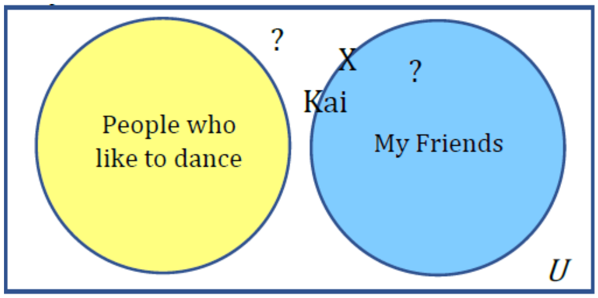


This argument is valid because we were able to show that the conclusion follows logically from the premises. But the argument is not sound because the premise that all water bottles are plastic is not true. There are many versions of glass and metal bottles that are evidence that the first premise is not true. This argument is valid but not sound.

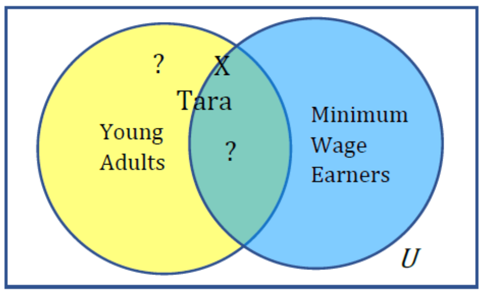
c. From the first premise we know that all firefighters lie inside the set of those who know CPR (firefighters are a subset of people who know CPR). From the second premise, we know that Jill is a member of the set of those who know CPR, but we do not have enough information to know whether she is also a member of set of firefighters.



Since we cannot determine which group Jill must be a part of, the argument is invalid. The statement that Jill is a firefighter does not follow logically from the premises that "all firefighters know CPR" and that "Jill knows CPR". Since the argument is not valid, it cannot be sound.

d. 

Because it said “none” we draw disjoint sets – one set for my friends and a second set for people who like to dance. The second premise tells us that Kai doesn't like to dance so they're not in the set of people who like to dance. However, we can't put Kai in the set of my friends either. They could be my friend, or someone I don't know who happens to not like dancing. Therefore, the conclusion is ***not valid***. And therefore, the arguement is also ***not sound***.

e. 

Because it said “some” we draw overlapping sets. The second premise tells us to put Tara in the set of young adults, but it doesn't tell us if she makes minimum wage or not. So, like the previous example we cannot determine which region she is in. She could make minimum wage, or she could also make more. Therefore, the conclusion is ***not valid*** and therefore, ***not sound***.

□

1.3.5 Exercises

1. Write whether the argument described is inductive or deductive:

a. The first five terms of the sequence were all odd, therefore the sixth term will also be odd.

b. My dog is afraid of loud noises. Today is the fourth of July, so there will be fireworks. My dog will be afraid today.

2. Find the next term of the sequences below.

a. 3, 6, 9, 12, 15, _________

b. 3, 6, 9, 15, 24, _________

Rewrite each of the following arguments in their “premise, premise, conclusion” form, and determine whether the argument is inductive or deductive. If the argument is inductive, determine its strength. If the argument is deductive, use sets to illustrate and determine the validity of the argument, and state whether the argument is valid and whether it is sound.

3. Since all cats are scared of vacuum cleaners and Max is a cat, Max must be scared of vacuum cleaners.

4. Every day for the last year, a plane flew over my house at 2 pm. Therefore, a plane will always fly over my house at 2pm.

5. Kiran collected data on the salaries of their friends. They found that female and nonbinary friends made less than male friends, so they concluded that women and nonbinary people make less than men.

6. Some of these kids are rude. Jimmy is one of these kids. Therefore, Jimmy is rude!

7. All bicycles have two wheels. My friend's Harley-Davidson has two wheels, so it must be a bicycle.

8. Since all chocolate contains milk and this bar is made of chocolate, it must contain milk.

9. All students drink a lot of caffeine. Brayer drinks a lot of caffeine, so he must be a student.

10. Over the course of a year, data was collected on the number of students visiting the cafeteria. On average, there were 15-35 students present in the cafeteria during the peak hours of the data. We can expect there to be between 15 and 35 students in the cafeteria if we go during the peak hours of the day.

11. If a person is on this reality show, they must be self-absorbed. Laura is not self-absorbed. Therefore, Laura cannot be on this reality show.

12. The first few terms of the sequence are 1, 3, 5, 7, 9. Therefore, the next term will be 11.

For each of the following, draw the appropriate illustration of sets (Subset, Disjoint or Overlapping). Then put an X to represent the subject of the conclusion or put two question marks to illustrate the subject could into two locations. Finally, state if the argument is valid and whether it is sound.

13. Premise: No apples are pears.

Premise: A Pink Lady is an apple.

Conclusion: Therefore, a Pink Lady is not a pear.

14. Premise: All children are young.

Premise: Tamika is young.

Conclusion: Therefore, Tamika is a child.

15. Premise: Some goats faint.

Premise: Fizzy faints.

Conclusion: Therefore, Fizzy is a goat.

16. Premise: All students who miss more than 25% of class time fail.

Premise: Claudia failed my class.

Conclusion: Claudia missed more than 25% of class time.

17. Premise: All students who miss more than 25% of class time fail.

Premise: Ethan missed more than 25% of class time.

Conclusion: Ethan failed.

18. Premise: All dogs eat apples.

Premise: Mary ate an apple.

Conclusion: Mary is a dog.

19. Premise: Some entering freshmen have to take a placement test.

Premise: Juan is an entering freshman.

Conclusion: Juan has to take a placement test.

20. Premise: No cats like peanut butter.

Premise: Bob does not like peanut butter.

Conclusion: Bob is a cat.

1.4 Logical Fallacies

Objectives: Section 1.4 Logical Fallacies.

Students will be able to:

- Identify common logical fallacies and their use in arguments

1.4.1 Logical Fallacies

In the last section we saw that logical arguments are invalid when the premises are not sufficient to guarantee the conclusion, and that even if an argument is valid it may be unsound if the premises are not true. There are other ways that a logical argument my by invalid or unsound. One of the more common ways this can occur is if the argument is a ***fallacy***.

A ***fallacy*** is a type of argument that appears valid but uses a logical error to persuade or deceive. Fallacious arguments are especially common in advertising and politics, so it is important as informed citizens to recognize when we are being presented with a fallacious argument and to not be persuaded by it.

1.4.2 Common Logical Fallacies

There are many logical fallacies, and some go by more than one name. Below we introduce a few of the more common fallacies that you will be asked to recognize by name, but there are many others.

1.4.3 Personal Attack (Ad hominem)

A ***personal attack*** argument attacks the person making the argument while ignoring the argument itself. A personal attack is not the same as an insult. Rather, a personal attack claims that there is something wrong with the person or group in order to cast doubt on their character and discredit their argument.

Example 1.4.1 "Jane says that whales aren't fish, but she's only in the second grade so she can't be right."

Here the argument is attacking Jane, not the validity of her claim, so this is a personal attack. □

Example 1.4.2 "Jane says that whales aren't fish, but everyone knows that they're really mammals. She's so stupid."

This certainly isn't very nice, but it is *not* a personal attack since a valid counterargument is made ("they really are mammals") along with a personal insult. □

Example 1.4.3 "Mr. Smith is a college dropout, so his stance on education reform cannot be trusted."

Here the argument uses the fact that Mr. Smith did not complete their college degree to discredit their ideas on education reform, so it is a personal attack. □

1.4.4 Appeal to Ignorance

An appeal to ignorance argument assumes something is true because it hasn't been proven false.

Example 1.4.4 "Nobody has proven that photo isn't of Bigfoot, so it must be Bigfoot."

This is an example of an appeal to ignorance since the fact that no one has been able to prove the picture of Bigfoot is false is being used as evidence that it is Bigfoot. □

1.4.5 Appeal to Authority

An ***appeal to authority*** argument attempts to use the authority of a person to prove a claim. An authority could be an expert such as a doctor or scholar, or someone who is admired like a celebrity or sports figure. While an authority can provide strength to an argument, problems can occur when the person's opinion is not shared by other experts, or when the authority is irrelevant to the claim.

Example 1.4.5 "A diet high in bacon can be healthy; Doctor Atkins said so."

Here, an appeal to a doctor's authority is used for the argument. This generally would provide strength to the argument, except that the opinion that eating a diet high in saturated fat runs counter to general medical opinion. More supporting evidence would be needed to justify this claim. □

Example 1.4.6 "Jennifer Hudson and Oprah lost weight with Weight Watchers, so their program must work."

In this example there is an appeal to the authority of celebrities. While their experience does provide evidence, it provides no more than any other person's experience would. □

1.4.6 False Dilemma

A ***false dilemma*** argument falsely frames an argument as an "either or" choice without allowing for additional options.

Example 1.4.7 "Either those lights in the sky were an airplane or aliens. There are no airplanes scheduled for tonight, so it must be aliens."

This argument is a false dilemma since it ignores the possibility that the lights could be something other than an airplane or aliens. □

1.4.7 Straw Man (or Straw Person)

A straw person argument involves misrepresenting the argument in an oversimplified, distorted and less favorable way to make it easier to attack.

Example 1.4.8 "Senator Khouri has proposed reducing military spending by 10%. Apparently, she wants to leave us defenseless against attacks by terrorists."

Here the arguer has represented a 10% funding cut as equivalent to leaving us defenseless, making it easier to attack Senator Khouri's position. □

1.4.8 Post Hoc

A ***post hoc*** argument claims that because two things happened sequentially, then the first must have *caused* the second.

Example 1.4.9 "Every morning the rooster crows just before dawn. It must be his crow that makes the sun rise."

Here the arguer is saying the rooster caused the sun to rise, but it is more likely that the sun rising caused the rooster to crow. □

Example 1.4.10 "Today I wore a red shirt and my football team won! I need to wear a red shirt every time they play to make sure they keep winning."

This person is saying their team won because they wore a red shirt. This type of superstition is quite common in sports even though we really know they are unrelated. □

Sometimes there may be more than one fallacy that seems reasonable. Consider this argument: "Emma Watson says she's a feminist, but she posed for these racy pictures. I'm a feminist and no self-respecting feminist would do that." Could this be ad hominem, saying that Emma Watson has no self-respect? Could it be appealing to authority because the person making the argument claims to be a feminist? Could it be a false dilemma because the argument assumes that a woman is either a feminist or not, with no gray area in between?

We have described just six of the many types of logical fallacies. Once you learn to recognize these you will also likely become aware of many others. There are many lists of logical fallacies online.

1.4.9 Exercises

Determine which type of fallacy each argument represents.

1. John Bardeen's work at the Advanced Institute for Physics has progressed so slowly that even his colleagues call him a plodder. Hence, it is prudent at present not to take seriously his current theory relating how strings constitute the smallest of subatomic particles.
2. You will tell the general manager that I made the right choice in dealing with that customer. After all, I'm the shift manager, so my decisions are always right.
3. It was his fault, Officer. You can tell by the kind of car I'm driving and by my clothes that I am a good citizen and would not lie. Look at the rattletrap he is driving and look at how he is dressed. You can't believe anything a dirty, longhaired hippie like that might tell you. Search his car; he probably has pot in it.
4. We can go to the amusement park or the library. The amusement park is too expensive, so we must go to the library.
5. Nearly all heroin addicts used marijuana before trying heroin. Clearly marijuana use leads to heroin addiction.
6. The oven was working fine until you started using it, so you must have broken it.
7. Old man Brown claims that he saw a flying saucer in his farm, but he never got beyond the fourth grade in school and can hardly read or write. He is completely ignorant of what scientists have written on the subject, so his report cannot possibly be true.
8. You should use Sparkle brand toothpaste since four out of five dentists recommend it.
9. She didn't say that I couldn't borrow her car, so I figured it was fine if I borrowed it for the weekend.
10. If you think that teens should be taught about contraceptive measures then you want to give kids license to have sex with no consequences.
11. You're either part of the solution or part of the problem.
12. No one can prove that God exists, therefore God does not exist.
13. You're clearly just too young to understand.
14. You should start off every morning with Champions brand cereal. It is what Michael Jordan eats, so you know that it must be very good for you.
15. No one on the council objected to the idea that he proposed, so everyone must think it is a good idea.
16. Why should we believe your testimony? You haven't had a steady job since 2003.
17. A huge percentage of diagnosed cases of autism came very soon after vaccinations for the measles. These vaccinations must be causing autism.
18. If you are against this war then you must hate America.
19. Just look at her face. How could anyone vote for that?

20. There are a number of fallacies that were not discussed in this section. Do an internet search for the following fallacies. Provide both a definition and at least one example.

a. Slippery Slope

b. Circular Reasoning

c. Appeal to Emotion

d. Red Herring

e. Whataboutism

1.5 Chapter 1 Review

Review Exercises

1. Identify which of the following is a proposition.

a. Airplanes are the safest form of travel.

b. Portland is a city in Oregon and Maine.

c. Is it raining outside?

d. Say what?

e. I'm going on vacation on Monday.

f. Everyone is addicted to technology.

2. Write the negation of each proposition.

a. I take public transportation to get to class.

b. I went to a movie on Friday.

c. I don't want to go golfing today.

d. I love watching basketball.

e. Breylynn's favorite color is green

f. Mirriam is a theater major.

3. For each situation, decide whether the "or" is most likely exclusive or inclusive.

a. I like watching soccer or basketball.

b. You should pack shorts or capris for our vacation.

c. We should take the train or bus to Portland to Eugene.

d. I would like to paint the room grey or blue.

e. The best reality TV show is Amazing Race or Big Brother.

4. Express in the form of *if p, then q.* Identify p and q

a. Doing homework helps increase your grade in class.

b. Riding public transportation will help you save money.

c. Squirrels, bury their food.

d. You will get sick if you eat too much candy.

e. Go to the doctor if you think you have the flu.

5. Translate each statement from symbolic notation into English sentences. Let A represent "I will buy an iPhone" and Let B represent "I learn how to use new technology fast".

a. Negation of A

b. Negation of B

c. A or B

d. A and not B

e. If B, then A

6. Complete the truth table for the following.

- Let A: I will buy an iPhone
- Let B: I learn how to use new technology fast.

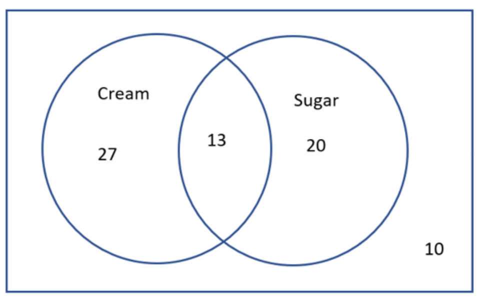
A	B	Not A	Not A and B	Not (Not A and B)
T	T			
T	F			
F	T			
F	F			

7. Complete the truth table for the following.

- Let A: I will buy an iPhone
- Let B: I learn how to use new technology fast.

A	B	If A, then B
T	T	
T	F	
F	T	
F	F	

8. The following Venn Diagram shows how 70 customers at the local coffee shop like their coffee.



a. How many people only like cream in their coffee?

b. How many people put sugar in their coffee?

c. How many people put sugar or cream in their coffee?

d. How many people don't like cream in their coffee?

9. In a group of 30 students at PCC, 21 have Biology this term, 12 have Math 105 and 7 are studying both Biology and Math 105.

a. Create a two-circle Venn diagram that summarizes the results of the survey.

b. How many students are not taking either course?

10. The following data was collected at a local coffee shop. 120 customers were asked if they like coffee and chocolate. The results are in the table below.

	Liked coffee	Disliked coffee	Total
Liked chocolate	23	37	60
Disliked chocolate	42	18	60

Create a two-circle Venn diagram that summarizes the results of the survey.

11. A 120 people were surveyed and it reveals the following results about how people learn about news events happening around the world.

- Twitter 35
- Websites 55
- Newspapers 10
- Newspaper and Website 5
- Newspaper and Twitte 4
- Website and Twitter 25
- All three sources 3

a. Create a three-circle Venn diagram that summarizes the results of the survey.

b. How many people get their news from Twitter or a Website?

c. How many people do not get their news from any of the three sources?

d. How many people use only Twitter?

e. How many people use Twitter or a Website but not a Newspaper?

For each argument, draw the appropriate illustration of sets (Subset, Disjoint or Overlapping).Then put an X to represent the subject of the conclusion. Alternatively, use two question marks to illustrate the subject could fit into two locations. Finally, state whether the conclusion is valid and whether it is sound.

12. All baseball parks have hotdogs.

Wrigley Field is a baseball park.

Conclusion: Wrigley Field has hotdogs.

13. All dogs to run in the park.

Henry does not like to run in the park.

Conclusion: Henry is not a dog.

14. Some students play sports.

Javeer plays soccer.

Conclusion: Javeer is a student.

Name the common Fallacy. Identify the Premise(s) and Conclusion.

15. Most people find out what's happening on Twitter or Facebook, so it is the most reliable source for news.

16. "Finding the Loch Ness Monster" has yet to provide evidence that Loch Ness exists, so all those sightings are obviously bogus.

17. Sampson bought a new car, and then he got a traffic ticket for speeding. Buying the new car must have caused him to speed.

Chapter 2

Financial Math

Dr. Muhammad Yunus and Microlending.

Muhammad Yunus[1] was born in 1940 in the village of Bathua to a Bengali Muslim family. He went to primary school in the city of Chittagong, Bangladesh and studied at Chittagong Collegiate School. In 1960 and 1961, Dr. Yunus completed his B.A. and M.A. in economics at Dhaka University. In 1965, he received a Fulbright scholarship to study in the United States and earned his Ph.D. in economics from Vanderbilt University. Then he taught economics at Middle Tennessee State University (Wikipedia Contributors, 2019).

After returning to Bangladesh he was appointed to the government's Planning Commission and then led the Chittagong Economics Department. In 1974, during a visit to the village of Jobra, Dr. Yunus met a woman who wove bamboo stools and was struggling to make ends meet. He loaned a total of US$27 of his own money to her and several women in the village. They repaid it in full and on time and were able to make a profit for themselves (Giridharas & Bradsher, 2006).

Dr. Yunus discovered that making very small loans could make a really big difference and empower people in poverty. He saw credit as a human right and invented the concept of microlending (Renton, 2008). Traditional banks did not want to risk making tiny loans to the poor and local lenders had excessive or abusive interest rates.

In December 1976, Yunus secured a loan from Janata Bank for a larger pilot program to lend money to villagers in Jobra and by 1983, the project became a full-fledged bank called Grameen Bank ("Village Bank"). In the late 1980s, Grameen started to diversify into fisheries, irrigation and other equity projects. From 1997 to 2007, the Village Phone project had brought cell phone ownership to 260,000 rural people in over 50,000 villages.

More than 94% of Grameen loans have gone to women, who are disproportionately affected by poverty. The model developed by Dr. Yunus inspired similar efforts in about 100 developing countries and in the United States, however, some businesses started taking advantage of people.

Dr. Yunus was awarded the Nobel Peace Prize in 2006, along with Grameen Bank, for their efforts in social business models and alleviating poverty. By July 2007, Grameen had issued US$6.38 billion to 7.4 million borrowers. To help ensure repayment, the bank uses a system of solidarity groups which are five women who

[1]Photo: "Muhammad Yunus (Cropped),"University of Salford Press Office is licensed under CC BY 2.0.

apply together and act as co-guarantors and support each other (Wikipedia Contributors, 2019).

Dr. Yunus and Grameen bank faced some accusations in 2011, which may have been politically motivated. He was cleared of any misconduct. Dr. Yunus currently chairs The Yunus Centre[2] in Dhaka, Bangladesh, a think tank for ideas related to social entrepreneurism, poverty alleviation and sustainability.

2.1 Introduction to Spreadsheets

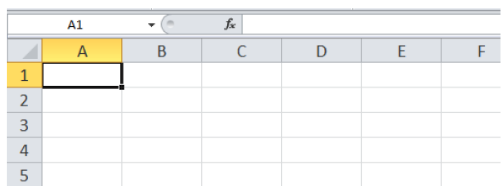
Objectives: Section 2.1 Introduction to Spreadsheets

Students will be able to:

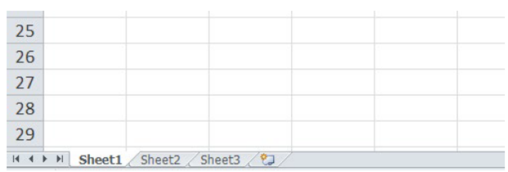
- Perform basic calculations on a spreadsheet
- Use cell references and the fill-down feature

A spreadsheet such as Google Sheets or Microsoft Excel, is a very useful tool for doing calculations and making complex tables. You can type in your own custom calculations or use the built-in formulas.

The rectangles within a spreadsheet are called ***cells***, and they can be referenced by their column letter and row number. The first cell in the upper left side highlighted below is A1. If we wanted to talk about the third column and the fifth row, that cell would be C5.



A spreadsheet file can contain many sheets. Look along the bottom to see if there is more than one sheet and make sure you are on the right sheet.

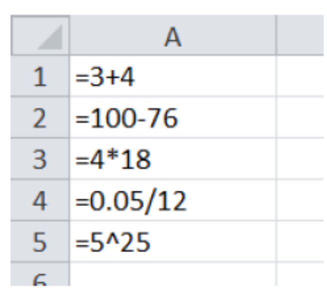


2.1.1 Basic Calculations

To do a calculation on a spreadsheet, type an equal sign before the operation. This lets the program know that you want it to calculate the result. When you press enter, you will see the result.

Example 2.1.1

1. To add 3 + 4, enter `=3+4`
2. To subtract 100-76, enter `=100-76`
3. To multiply 4 times 18, enter `=4*18`
4. To divide 0.05 by 12, enter `=0.05/12`
5. To calculate 5^{25}, enter `=5^25`



[2] https://www.muhammadyunus.org/

□

Note that the asterisk (∗) is used for multiplication. Spreadsheets don't recognize parentheses as indicators of multiplication like calculators do, so even if you have parentheses for the order of operations, the asterisk is also needed.

You can make more complicated mathematical expressions using parentheses and other operations. To ***edit a cell*** click on the editing box at the top, or double click on the cell to edit it directly.

Example 2.1.2 Your bill at a restaurant is \$35.75 and you want to leave an 18% tip. How much would you add to the bill?

Solution. To work with a percentage, we need to convert it into a decimal first, and then multiply it by the base amount.

In a spreadsheet we would type

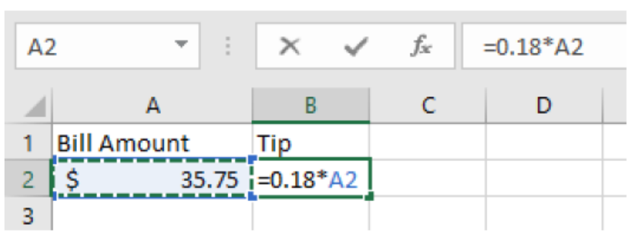
```
=0.18*35.75
```

and get a result of \$6.44, ***rounded to the nearest cent***. You would leave a tip of \$6.44. □

2.1.2 Cell References

One of the powerful things about spreadsheets is using a cell reference, such as C5 in a calculation. When you use a ***cell reference***, the values will automatically update if any of the referenced values change.

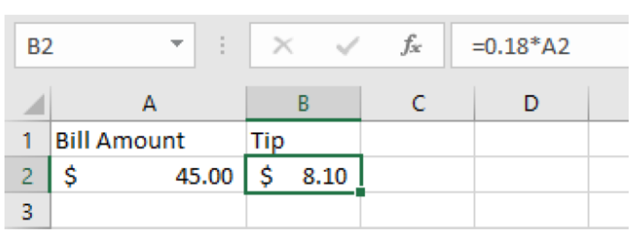
Let's make a spreadsheet for the percentage tip example above. We calculated an 18% tip on a bill of \$35.75. We might want to tip 18% in general, but our bill will change values. We labeled the first column Bill Amount and the second column Tip. The amount of \$35.75 is entered in cell A2. Then when we write our formula in B2, we want to calculate 18% of A2. That way if the number in A2 changes, our tip will automatically update.



The formula

```
=0.18*A2
```

is entered in B2 which gives a result of \$6.44 when you hit enter.



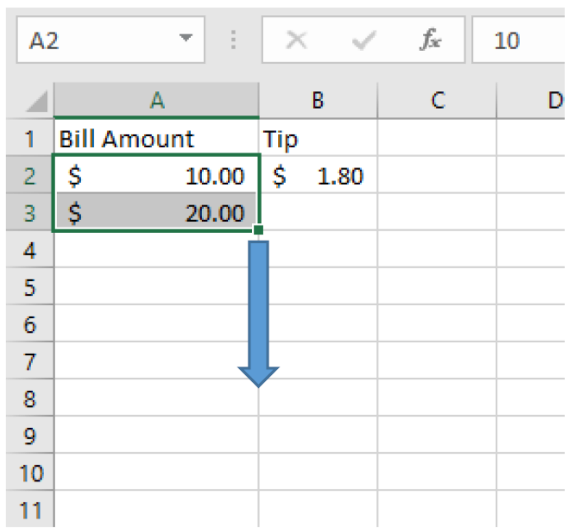
When the bill amount is changed, the tip is recalculated.

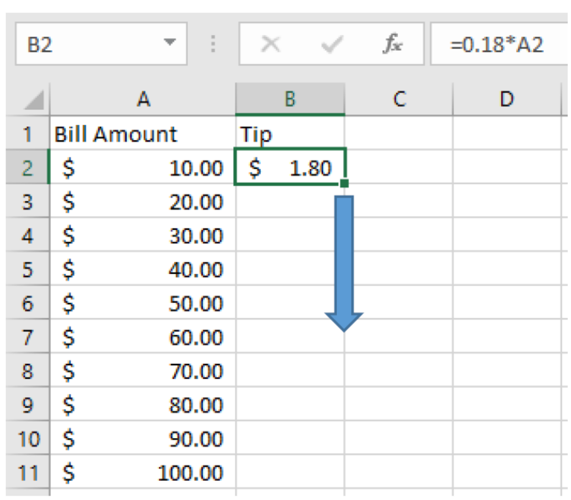
2.1.3 Cell Formatting

We can also format cells A1 and B1 to show dollar signs by clicking on the dollar sign in the number formatting menu.

2.1.4 Fill-Down Feature

The ***fill-down feature*** is very useful for making tables. This allows us to copy values or formulas to save time. Let's make a tipping reference table with values from \$10, to\$100, in increments of \$10. First, we will enter two values in column A to establish the pattern. Then select those two cells and you will see a small square in the lower right corner. Drag that square down until you get to \$100.





Next, we can drag our formula down and the cell reference will change to each row number automatically.

Here are the formulas with the row numbers updated:

B2 =0.18*A2

	A	B	C
1	Bill Amount	Tip	
2	10	=0.18*A2	
3	20	=0.18*A3	
4	30	=0.18*A4	
5	40	=0.18*A5	
6	50	=0.18*A6	
7	60	=0.18*A7	
8	70	=0.18*A8	
9	80	=0.18*A9	
10	90	=0.18*A10	
11	100	=0.18*A11	

Here is our completed table with the calculations:

B11 =0.18*A11

	A	B	C	D
1	Bill Amount	Tip		
2	$ 10.00	$ 1.80		
3	$ 20.00	$ 3.60		
4	$ 30.00	$ 5.40		
5	$ 40.00	$ 7.20		
6	$ 50.00	$ 9.00		
7	$ 60.00	$ 10.80		
8	$ 70.00	$ 12.60		
9	$ 80.00	$ 14.40		
10	$ 90.00	$ 16.20		
11	$ 100.00	$ 18.00		

2.1.5 Formulas

Spreadsheets have many useful built-in formulas. We will introduce some of the financial formulas in this chapter. Here are some of the formulas we will use:

1. `=FV` to calculate the future values of an investment
2. `=PV` to calculate the deposit needed for a desired future balance
3. `=PMT` to calculate a loan or savings plan payment
4. `=EFFECT` to calculate the effective rate of an account and compare accounts

In the rest of this chapter we will use spreadsheets and formulas to calculate the future values, interest paid or earned and monthly payments.

2.1.6 Exercises

Use a spreadsheet to compute the following.

1. Convert $4/7$ to a decimal

2. Convert 16% to a decimal

3. Add 8 and 19

4. Find the difference of 230 and 78

5. Multiply 12 and 9

6. Divide 0.09 by 52

7. Calculate 8^3

8. Your bill at a restaurant is $55.75 and you want to leave a 20% tip. How much would you add to the bill?

9. You leave a tip for $7.50 for a bill at a restaurant that is $44.50. What percent tip did you leave?

10. In Column A use the fill down feature to build a spreadsheet starting with $5 and ending at $125, in increments of $5. In Column B write a formula with a cell reference to calculate a 15.5% tip on the amount in Column A. Use the fill down feature to complete your table.

11. Imagine a certain savings account started out with a balance of $5250.00 on day-one, and today has a current balance of $5780.23

a. Exactly how much more money does the account have today, compared with day-one?

b. Rounding to the nearest tenth of a percent: By what percentage amount has the account balance grown?

c. If instead, the bank balance today was exactly double the starting balance, then by what exact percentage amount would the bank balance have grown?

d. If the bank balance today had instead grown by 15.5% since day-one, then what would be the exact amount of today's balance?

12. Imagine that at the start of a certain month, you will make an opening deposit of $500 into a savings account, and you will then leave the account alone (meaning you will make no further deposits or withdrawals). Also, for this account: Every month after the opening deposit, the amount in the account will grow to be 101% of its previous month's balance.

a. Use a spreadsheet to enter 500 in cell A1. Using a formula and a cell-reference: Compute in cell A2, the amount in the account after one month has passed. Then using the fill down feature, continue the pattern for another eleven full months (you should end at cell A13). Format all the cells to show dollar signs. What is the amount in the account after one year?

b. Now continue the pattern in column A of your spreadsheet to extend for a second full year (you should end at cell A25). What is the amount in the account after two years?

c. What overall percentage growth occurred in the account between the opening deposit and one year later? (Compute using a formula and cell references)

d. What overall percentage growth occurred in the account between the end of year one, and the end of year two? (Compute using a formula and cell references)

e. (Challenge) The annual percentage growth that you found in part (d) for the second year, should be identical to the annual percentage growth that you found in part (c) for the first year. Can you mathematically explain why this is true? Do you think this pattern of identical overall annual percentage growth would continue, if you extend the pattern for even more years?

13. Imagine that at the start of a certain year, you will deposit $1000.00 into a savings account, and then you will leave the account alone. Each year after the opening deposit, the amount in the account will grow to be 103% of its previous year's balance.

a. After two years, the account balance will have experienced two growth amounts of 103%. You can find this account balance amount here, with the spreadsheet computation `= 1000 * (103%) * (103%)`. Perform this computation in a spreadsheet and write the balance that you find.

b. Now enter the spreadsheet computation `= 1000 * (103%)^2`. Notice that the result here, which involves using a power, gives the same answer as you found in part (a). Comparing the two spreadsheet computations: Explain why they give the same result.

c. Using the pattern in part (b) above, and carefully choosing the power: Compute the balance that will be in the account fifteen full years after the account was originally opened. (Round to the nearest cent)

d. (Challenge) Make a spreadsheet that shows the account balance each individual year for 30 years. From the date of the opening deposit: What minimum number of full years will you have to wait, until the balance finally exceeds twice its opening deposit amount? (Use cell references, the fill down feature, and dollar formatting)

e. (Challenge) Imagine the opening balance of the account was \$5000.00 instead of \$1000.00 (and everything else about the account stays the same). Make a similar spreadsheet as you did in part (d), and using this spreadsheet, find the minimum number of full years you will have to wait this time, until the balance finally exceeds twice its opening deposit amount. How does this answer compare with your answer in part (d)? Do you think your answer would be the same here, for any positive opening balance you may choose for the account?

2.2 Simple and Compound Interest

Objectives: Section 2.2 Simple and Compound Interest

Students will be able to:

- Use spreadsheet functions and/or mathematical formulas to calculate simple, compound, and continuously compounded interest
- Understand the difference between simple and compound interest
- Use a spreadsheet to calculate the effective rate and compare accounts
- Use a spreadsheet and/or formula to calculate the present value needed to reach a desired future value

Note: Spreadsheets are emphasized in this chapter, but the formulas are also presented so you can understand what the spreadsheet is doing. Be sure to check with your instructor for which method to use.

Working with money is a very important skill for everyday life. While balancing a checkbook or calculating our monthly expenditures on espresso requires only arithmetic, when we start saving money, planning for retirement, or need a loan, we need more mathematics and tools. In this section we will calculate and compare simple and compound interest.

2.2.1 Simple Interest

Calculating interest starts with the ***principal***, P, or the beginning amount in your account. This is also called the ***present value***. This could be a starting investment, or the starting amount of a loan. The ***interest rate***, I, in its most simple form, is a percentage 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 be 5% of 100. It is very important to remember to change the interest ***rate***, r, of 5% into a decimal by moving the decimal two places to the left.

$$(0.05)\$100 = \$5$$

The total amount you would repay is called the ***future value*** and would be \$105, the original principal plus the interest.

$$\$100 + \$5 = \$105$$

Here are the formulas to represent the calculations we just did.

Simple One-time Interest.

$$I = Pr$$

$$A = P + I \text{ or } A = P + Pr$$

I is the interest

P is the principal, starting amount or ***present value***

r is the interest rate in decimal form

A is the end amount: principal plus interest. This is also called the ***future value***

Example 2.2.1 A friend asks to borrow $300 and agrees to repay it in 30 days with 3% simple interest. How much interest will you earn?

Solution. $P = \$300$ the principal or present value

$r = 0.03$ 3% rate

Using the formula,

$$\begin{aligned} I &= Pr \\ &= \$300(0.03) \\ &= \$9 \end{aligned}$$

To calculate this in a spreadsheet, you would enter

```
=300*0.03
```

and get a result of $9. You will earn $9 in interest when your friend pays you back. □

One-time simple interest is only common for extremely short-term or informal loans. For longer term loans or investments, it is common for interest to be paid on a daily, monthly, quarterly, or annual basis. In that case, interest would be earned regularly. Bonds are an example of this type of investment. Bonds are issued by the federal, state or local governments to cover their expenses.

Example 2.2.2 Suppose your city is building a new park, and issues bonds to raise the money to build it. You buy a $1,000 bond that pays 5% simple interest annually and matures in 5 years. How much interest will you earn? What is the future value of the bond?

Solution. Each year, you would earn 5% interest so over the course of five years, you would earn:

$$\$1000(0.05)(5) = \$250$$

When the bond matures, you would receive back the $1,000 you originally paid and the $250 in interest, so we could also put that into a single calculation:

$$\$1000 + \$1000(0.05)(5) = \$1250$$

Using a spreadsheet, you would enter

```
=1000+1000*0.05
```

and get a result of $1,250. The future value of the bond is $1,250. □

We can generalize this idea of simple interest over time.

Simple Interest over Time.

$$I = Prt$$

$$A = P + I \text{ or } A = P + Prt$$

I is the interest

P is the principal, starting amount, or present value

r is the interest rate in decimal form

t is time, where the increment of time (years, months, etc.) matches the time period for the interest rate

A is the end amount, principal plus interest, or future value

2.2.2 APR – Annual Percentage Rate

Interest rates are usually stated as an annual percentage rate (APR) – the total interest that will be paid in the year. If not stated otherwise, assume that the interest rate is an annual rate or APR. If the interest is paid in smaller time increments, the APR will be divided by the number of time periods.

Note: The Federal Truth in Lending Act requires that every consumer be given the true APR which includes the interest and any fees included.

For example, a 6% APR paid monthly, would be divided by 12, because you would get one twelfth of the rate per month, which is half a percent per month.

$$\frac{0.06}{12} = 0.005$$

A 4% annual rate paid quarterly, would be divided by 4 to get 1% per quarter.

$$\frac{0.04}{4} = 0.01$$

Here is an example of a semi-annual rate.

Example 2.2.3 Suppose you buy a \$1,000 federal bond with a 4% annual simple interest rate, paid semi-annually, with a maturity in 4 years. How much interest will you earn? What will be the future value of the bond?

Solution. $P = \$1000$ the principal

$r = \frac{0.04}{2} = 0.02$ interest is being paid semi-annually (twice a year), so the 4% interest will be divided into two 2% payments.

$t = 8$ 4 years, compounded twice a year so $t = 4 \cdot 2 = 8$ half-years

$$\begin{aligned} I &= PRT \\ &= \$1000(0.02)(8) \\ &= \$160 \end{aligned}$$

You will earn \$160 in interest over the four years. The future value of the loan is

$$\begin{aligned} A &= P + I \\ &= \$1000 + \$160 \\ &= \$1160 \end{aligned}$$

We could also use a spreadsheet to do this calculation and enter:

```
=1000+1000*(0.04/2)*(4*2)
```

which also gives $1,160. The future value of the bond is $1,160. Remember that spreadsheets don't interpret parentheses as multiplication. We need the asterisks as well as the parentheses. □

2.2.3 Compound Interest

In a standard bank account, any interest we earn is automatically added to our balance, and we earn interest on that interest. This reinvestment of interest is called ***compounding***. We will develop the mathematical formula for compound interest and then show the equivalent spreadsheet function.

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 $\frac{0.03}{12} = 0.0025$ per month.

In the first month,

$$\begin{aligned} P &= \$1000 \\ r &= 0.0025 \text{ (which is 0.25\%)} \\ I &= \$1000(0.0025) = \$2.60 \\ A &= \$1000 + \$2.5 = \$1002.50 \end{aligned}$$

In the first month, we will earn $2.50 in interest, raising our account balance to $1002.50.

In the second month,

$$\begin{aligned} P &= \$1002.50 \\ I &= \$1002.50(0.0025) = \$2.51 \text{ (rounded)} \\ A &= \$1002.50 + \$2.51 = \$1005.01 \end{aligned}$$

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 gives us.

Calculating out a few more months in a table or a spreadsheet we have:

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	2.56	1030.42

To find an equation to represent this, we will go through a few months to see the pattern:

Initial Amount: $P = \$1000$

1^{st} Month $A = 1.0025(\$1000)$

2^{nd} Month $A = 1.0025(1.0025(\$1000)) = 1.0025^2(\$1000)$

3^{rd} Month $A = 1.0025(1.0025^2(\$1000)) = 1.0025^3(\$1000)$

4^{th} Month $A = 1.0025(1.0025^3(\$1000)) = 1.0025^4(\$1000)$

Observing a pattern, we could conclude

n^{th} Month $A = 1.0025^n(\$1000)$

Notice that the \$1000 in the equation was P, the starting amount. We found 1.0025 by adding one to the interest rate divided by 12, since we were compounding 12 times per year. Generalizing our result, we could write

Compound Interest.

$$A = P\left(1+\frac{r}{n}\right)^{nt} \text{ or } P = \frac{A}{\left(1+\frac{r}{n}\right)^{nt}}$$

A is the future value balance in the account after n years

P is the principal or present value

r is the annual interest rate in decimal form

n is the number of compounding periods in one year

t is the number of years

If the compounding is done annually (once a year), $n = 1$.

If the compounding is done quarterly, $n = 4$.

If the compounding is done monthly, $n = 12$.

If the compounding is done weekly, $n = 52$.

If the compounding is done daily, $n = 365$.

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.

2.2.4 The Future Value Spreadsheet Formula

The compound interest formula is built into spreadsheets and is called the future value formula.

Future Value Spreadsheet Formula.

```
=FV(rate per period, number of periods, payment amount, present value)
```

rate per period is the interest rate per compounding period, r/n

number of periods is the total number of periods, $n * t$

payment amount is the amount of regular payments. If none, enter 0

present value is the amount deposited or principal, P

We will use the payment amount in a future section, for now that will be 0. There is also an optional input at the end to specify making payments at the beginning or end of the period, but we will not use it in this book.

Now let's look at an example and calculate the compound interest using the spreadsheet and the mathematical formula.

Example 2.2.4 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% APR, compounded monthly. How much will you have in the account after 20 years?

Solution. $P = \$3000$ the initial deposit

$r = 0.06$ 6% annual rate

$n = 12$ 12 months in 1 year

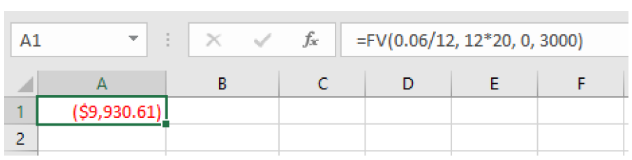
$t = 20$ since we're looking for how much we'll have after 20 years

To use a spreadsheet, we will use the formula like this

```
=FV(rate per period, number of periods, payment amount, present value)
```

```
=FV(0.06/12, 12 / 20, 0, 3000)
```

and get a result of \$9,930.61, rounded to the nearest cent.



Note that the output of the formula gives the answer with the opposite sign as the principal and payments. A negative number may be denoted with a negative sign or with the color red or parentheses. The signs may be used in accounting, but we will ignore them in this book.

To use a formula, we are looking for the future value, so we use the formula solved for A:

$$\begin{aligned} A &= 3000\left(1+\frac{0.06}{12}\right)^{12\cdot 20} \\ &= \$9930.61 \end{aligned}$$

To use a calculator, you would enter the formulas including parentheses around any inside operations. You would enter

```
3000(1+(0.06/12))^(12*20)=$9930.61
```

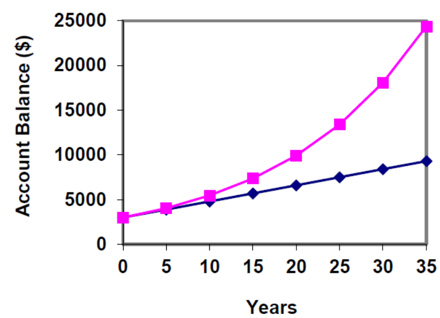
□

2.2.5 Comparing Simple and Compound Interest

Let us compare the amount of money earned from compounding in the previous example against the amount you would earn from simple interest. From the table and graph below we can see that 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.

Table 2.2.5

Year	Simple Interest ($15 per month)	Compound Interest (6% compounded monthly or 0.5% each month
0	$3000	$3000
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



2.2.6 Finding the Principal, or Present Value

When we know the amount of money we want to have in the future, we can use the formula that is solved for P. It requires a little algebra to divide both sides of the formula by the quantity that was multiplied by P. There is also a spreadsheet formula which we will introduce now, and then do an example using both methods.

2.2.7 The Present Value Spreadsheet Formula

The present value spreadsheet formula will calculate how much you need to deposit in the present to get a specified future value.

Present Value Spreadsheet Formula.

```
=PV(rate per period, number of periods, payment amount, future value)
```

rate per period is the interest rate per compounding period, r/n

number of periods is the total number of periods, $n * t$

payment amount is the amount of regular payments. If none, enter 0

future value is the amount desired in the future, A

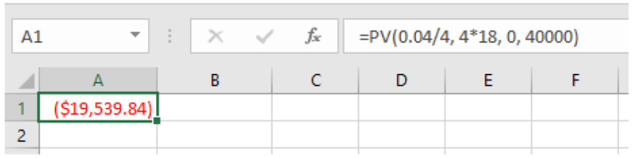
Example 2.2.6 You know that you will need $40,000 for your child's education in 18 years. If your account earns 4% APR compounded quarterly, how much would you need to deposit now to reach your goal?

Solution. We are looking for what we need to deposit now so we will use the present value formula. We type the formula and inputs the same way we used the future value formula.

```
=PV(rate per period, number of periods, payment amount, future value)
```

```
=PV(0.04/4, 4*18, 0, 40000)
```

which gives a value of $19,539.84. You would need to deposit $19,539.84 now and keep the same interest rate to have $40,000 in 18 years.



Note that we cannot enter commas in numbers in a spreadsheet. Commas are used to separate the input values, so we would not get the same answer if we put in $40,000 for an input.

To use the mathematical formula, we use the one that is solved for P.

$r = 0.04$ 4% APR

$n = 4$ 4 quarters in 1 year

$t = 18$ Since we know the balance in 18 years

$A = \$40,000$ The amount we have in 18 years

In this case, we're going to have to set up the equation, and solve for P.

$$\begin{aligned} P &= \frac{4000}{\left(1 + \frac{0.04}{4}\right)^{4 \cdot 18}} \\ &= 19539.84 \end{aligned}$$

You would need to deposit $19,539.84 now to have $40,000 in 18 years. □

2.2.8 Continuously Compounded Interest

In many bank accounts your interest is compounded continuously, or at each moment in time. The number of times per year, n, is infinite. As n approaches infinity the compound interest formula changes to the continuously compounded interest formula.

Continuously Compounded Interest.

$$A = Pe^{rt} \text{ or } P = \frac{A}{e^{rt}}$$

A is the future value or desired balance in the account

P is the principal or present value

r is the annual interest rate in decimal form

t is the number of years

e is an irrational number that is approximately 2.718281828... Find e on your calculator to use this formula

To calculate this on a spreadsheet we use the `=EXP` function. The spreadsheet formulas are

`=Principal*EXP(r*t) or =A/EXP(r*t)`

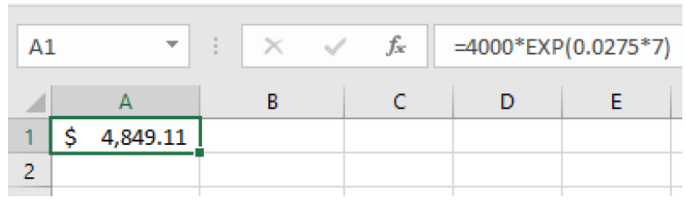
Example 2.2.7 You deposit $4000 in an account that earns 2.75% APR compounded continuously. How much will you have after 7 years? How much interest did you earn? What percentage of the final balance is interest?

Solution.

To use a spreadsheet, we look at the formula solved for A, the future value. We enter

```
=4000*EXP(0.0275*7)
```

and get a result of $4849.11.



To use the formula, we have:

$A = Pe^{rt}$

$P = 4000$ Amount invested

$r = 0.0275$ Interest rate

$t = 7$ Number of years

$$\begin{aligned} A &= 4000e^{(0.0275 \cdot 7)} \\ &= \$4849.11 \end{aligned}$$

After 7 years your account would be worth $4,849.11. Next, we will calculate the amount of interest earned and the percentage. □

2.2.9 Finding the Amount of Interest Earned and the Percentage

In the previous example we also want to know how much interest was earned and what percentage of the final balance is from interest. The future value of the investment is $4,849.11. Now to figure out how much of that was interest, we need to subtract the amount initially deposited.

Example 2.2.8 Continued: To find the total amount of interest earned, we subtract the principal from the total balance.

$$\$4849.11 - \$4000 = \$849.11$$

The spreadsheet calculation is

```
=4849.11-4000
```

which gives a result of $849.11. You would earn $849.11 in interest.

To find the percentage that is interest, divide the amount of interest by the total amount.

$$\frac{\$849.11}{4849.11} = 0.1751 \text{ or } 17.5\%$$

The spreadsheet calculation is the same:

```
=849.11/4849.11
```

which is 0.1751 or 17.5%. This tells us that after 7 years, 17.5% of the account was earned as interest. □

2.2.10 Effective Rate

If you are shopping around for different investments, you might need to compare different rates that have different compounding periods. If the rate and period are different, it's hard to know which account will give the better result. There is a spreadsheet formula called `=EFFECT` which will allow us to compare accounts. This is also sometimes called the annual percentage yield, or APY.

Effective Rate Formula.

`=EFFECT(stated rate, number of compounding periods)`

stated rate is the interest rate given (APR)

number of compounding periods is the number of times the account is compounded per year, n

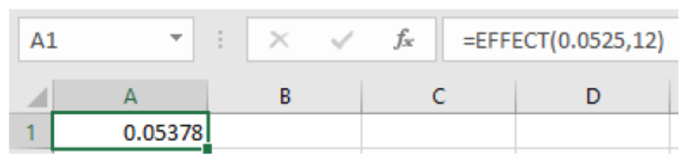
Example 2.2.9 You are comparing an account that pays 5.25% APR compounded monthly, with an account that pays 5% APR compounded daily. Which account will earn you more interest?

Solution. It is hard to tell whether the higher interest rate will be better or the higher compounding rate in this case. We will find the effective rate of both accounts.

For the 5.25% APR account compounded monthly:

`=EFFECT(0.0525,12)`

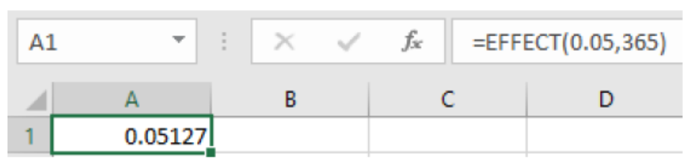
which gives 0.05378 or 5.38%



For the 5% APR account compounded daily:

`=EFFECT(0.05,365)`

which gives 0.05127 or 5.13%



Now we can compare the effective rates of 5.38% and 5.13% and see that the account with the higher interest rate will earn more interest in this case. This is not always true, so we will show another example. □

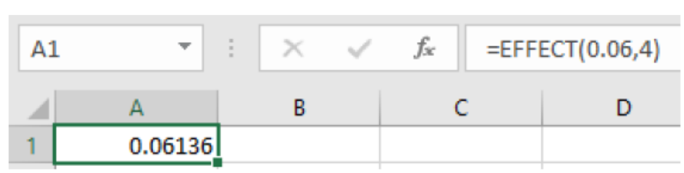
Example 2.2.10 Find the effective rates to compare an account that earns 6% APR compounded quarterly with an account that earns 5.975% APR compounded daily. Which one would you choose?

Solution. Using the effective rate formula for each, we have:

For the 6% APR account compounded quarterly:

`=EFFECT(0.06,4)`

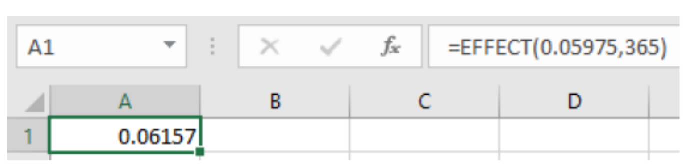
which gives 0.06136 or 6.14%



For the 5.975% APR account compounded daily:

`=EFFECT(0.05975,365)`

which gives 0.06157 or 6.16%



The account that was compounded more often has a slightly higher rate in this case. □

2.2.11 Exercises

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. You loan your friend \$100. They agree to pay an annual interest rate of 3%, simple interest. Six months later they repay that loan.

a. How much did they pay you?

b. How much was interest?

3. Consider a simple interest loan of \$200 with an annual interest rate of 6%. If that loan is paid off 1 year and 3 months later, how much was repaid?

4. You deposit \$1,000 in an account that earns simple interest. The annual interest rate is 2.5%.

a. How much interest will you earn in 5 years?

b. How much will you have in the account in 5 years?

5. Consider an investment of \$20000 with an annual interest rate of 5%.

a. If that investment is earning simple interest, how much will the investment be worth in 10 years?

b. If that investment is getting annually compounding interest, how much will the investment be worth in 10 years?

6. Nico invests \$4,500 into an account that has an annual interest rate of 8.5%. The interest is compounding monthly. Twenty years later what is the account balance?

7. How much will \$1,000 deposited in an account earning 7% APR compounded weekly be worth in 20 years?

8. Suppose you obtain a \$3,000 Certificate of Deposit (CD) with a 3% APR, paid quarterly, with maturity in 5 years.

a. What is the future value of the CD in 5 years?

b. How much interest will you earn?

c. What percent of the balance is interest?

9. You deposit \$300 in an account earning 5% APR compounded annually. How much will you have in the account in 10 years?

a. How much will you have in the account in 10 years?

b. How much interest will you earn?

c. What percent of the balance is interest?

10. You deposit \$2,000 in an account earning 3% APR compounded monthly.

a. How much will you have in the account in 20 years?

b. How much interest will you earn?

c. What percent of the balance is interest?

d. What percent of the balance is the principal?

11. You deposit \$10,000 in an account earning 4% APR compounded weekly.

a. How much will you have in the account in 25 years?

b. How much interest will you earn?

c. What percent of the balance is interest?

d. What percent of the balance is the principal?

12. 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% APR compounded monthly.

13. 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% APR compounded quarterly.

14. Breylan invests $1,200 in an account that earns 4.6% APR compounded quarterly and Angad invests the same amount in an account that earns 4.55% APR compounded weekly.

a. What will their balances be after 15 years?

b. What will their balances be after 30 years?

c. What is the effective rate for each account?

15. Bill invests $6,700 in a savings account that compounds interest monthly at 3.75% APR. Ted invests $6,500 in a savings account that compound interest annually at 3.8% APR.

a. Find the effective rate for each account.

b. Who will have the higher accumulated balance after 5 years?

16. Bassel is comparing two accounts where one pays 3.45% APR quarterly and the second pays 3.4% APR daily.

a. What is the effect rate for each?

b. If he has $5,000 to deposit how much will the balance be in 10 years?

17. You deposit $2,500 into an account earning 4% APR compounded continuously.

a. How much will you have in the account in 10 years?

b. How much total interest will you earn?

c. What percent of the balance is interest?

18. You deposit $1,000 into an account earning 5.75% APR compounded continuously.

a. How much will you have in the account in 15 years?

b. How much total interest will you earn?

c. What percent of the balance is interest?

19. You deposit $5,000 in an account earning 4.5% APR compounded continuously.

a. How much will you have in the account in 5 years?

b. How much total interest will you earn?

c. What percent of the balance is interest?

20. You deposit $10,000 in an account that earns 5.5% APR compounded continuously and your friend deposits $10,000 in an account that earns 5.5% APR compounded annually.

a. How much more will you have in the account in 10 years?

b. How much more interest did you earn in the 10 years?

2.3 Savings Plans

Objectives: Savings Plans 2.3

Students will be able to:

- Use a spreadsheet and/or formula to calculate the future value and interest earned on savings plans
- Use a spreadsheet and/or formula to calculate payment amounts for savings plans
- Analyze and compare lump sum and regular payment savings plans

For most of us it is not practical to deposit a large sum of money in the bank. Instead, we save by depositing smaller amounts of money regularly. We might save in an IRA or 401-K for retirement. We might also save for a down payment on a car or house, or in a college savings plan for our children.

Just like the last section, we will emphasize spreadsheets but calculate each example with the formulas as well. Check with your instructor for which way you should do your problems.

2.3.1 Savings Plan Formulas

To make calculations for savings plans using a spreadsheet, we can use the `=FV` formula we have already used. This time for regular payments we will use the field for payment amount. If we are not making an initial deposit, the present value will be zero.

Here is the future value formula again:

Future Value Spreadsheet Formula.

```
=FV(rate per period, number of periods, payment amount, present value)
```

rate per period is the interest rate per compounding period, r/n

number of periods is the total number of periods, $n * t$

payment amount is the amount of regular payments

present value is the initial principal. If none, enter 0

The mathematical formulas are shown below. If you want to know how we got the formula, it is derived at the end of the chapter.

Savings Plan Formulas.

$$A = \frac{d\left[\left(1+\frac{r}{n}\right)^{nt}-1\right]}{\left(\frac{r}{n}\right)} \text{ or } d = \frac{A\left(\frac{r}{n}\right)}{\left[\left(1+\frac{r}{n}\right)^{nt}-1\right]}$$

A is the balance in the account after n years (future value)

d is the regular deposit (or payment amount each month, quarter, year, etc.)

r is the annual interest rate in decimal form

n is the number of compounding periods in one year

t is the number of years

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 make your deposits every year, use yearly compounding, $n = 1$.

If you make your deposits every quarter, use quarterly compounding, $n = 4$.

If you make your deposits every month, use monthly compounding, $n = 12$, etc.

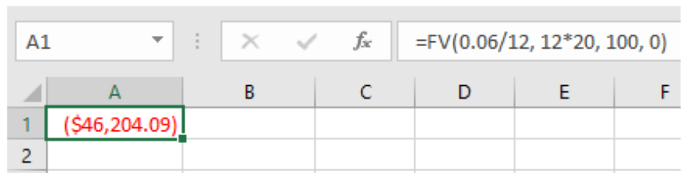
To see how both of these methods work, let's look at an example.

Example 2.3.1 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% APR, how much will you have in the account after 20 years? How much will you have earned in interest? What percentage of the balance is interest?

Solution. To use a spreadsheet, we will use `=FV` because we want to know the balance in the future. We enter 100 for the payment amount and 0 for the present value:

```
=FV(0.06/12, 12 / 20, 100, 0)
```

which gives a result of \$46,204.09.



Remember that the output of the formula gives the answer with the opposite sign as the principal and payments. For our purposes we will ignore the signs. To use the formula, we use the one solved for A, since we want to know the final amount.

$d = \$100$ the monthly deposit

$r = 0.06$ 6% annual rate

$n = 12$ since we're doing monthly deposits, we'll compound monthly

$t = 20$ we want the amount after 20 years

Putting this into the equation we have:

$$\begin{aligned} A &= \frac{100\left[\left(1+\frac{0.06}{12}\right)^{12\cdot 20}-1\right]}{\left(\frac{0.06}{12}\right)} \\ &= \frac{100[(1.005)^{240}-1]}{(0.005)} \\ &= \$46.204.09 \end{aligned}$$

With U.S. dollars we round to the nearest cent. The account will grow to \$46,204.09 after 20 years.

To find the ***amount of interest earned***, calculate the total of all your deposits.

$$\$100(20)(12) = \$24,000$$

The difference between the total amount and the deposits is the interest earned.

$$\$46,204.09 - \$24,000 = \$22,204.09$$

The total amount of interest you earned was \$22,204.09.

To find the ***percentage of the balance that is interest*** we will divide the interest by the total balance.

$$\frac{\$22,204.09}{\$46,204.09}$$

or 48.1%. After 20 years 48.1% of the balance is from interest. □

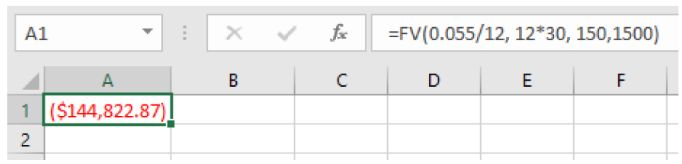
Now here's an example with an initial deposit and monthly deposits. We can do this with the spreadsheet formula.

Example 2.3.2 You want to jumpstart your saving by depositing $1500 from your tax return and then deposit $150 every month into an account that earns 5.5% APR compounded monthly. How much will you have in the account after 30 years?

Solution. Using the spreadsheet formula, we can enter an initial deposit and a monthly payment. We enter

```
=FV(0.055/12, 12*30, 150,1500)
```

and get a result of $144,822.87.



□

2.3.2 Finding Payment Amounts Usings Spreadsheets and Formulas

Another important thing we can calculate is how much we need to save in each period to have a specified amount in the future. Say you want to achieve a certain amount for retirement or for your kids' college.

The mathematical formula for this is the one solved for d, the payment amount, above. There is a new spreadsheet formula to calculate payments, =PMT, that we will introduce now.

Payment Spreadsheet Formula.

```
=PMT(rate per period, number of periods, present value, future value)
```

rate per period is the interest rate per compounding period, r/n

number of periods is the total number of periods, $n * t$

present value is the amount deposited or principal, P

future value is the amount you want in the future, A

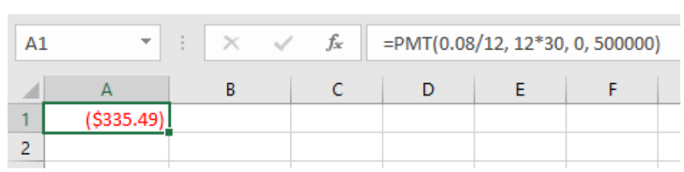
Here is an example of a retirement goal calculated with a spreadsheet and the formula.

Example 2.3.3 You want to have half a million dollars in your account when you retire in 30 years. Your retirement account earns 8% APR. How much do you need to deposit each month to meet your retirement goal?

Solution. To calculate this with a spreadsheet, we will use the `=PMT` function and enter 0 for the present value and $500,000 for the future value. We cannot enter commas within the numbers however, because spreadsheets use commas to separate the inputs. We enter:

```
=PMT(0.08/12, 12*30, 0, 500000)
```

and get a result of $335.49.



To see how this works with the formulas, we use the one solved for d, the regular deposit amount.

$r = 0.08$ 8% annual rate

$n = 12$ since we're depositing monthly

$t = 30$ 30 years

$A = \$500,000$ The amount we want to have in 30 years

$$\begin{aligned} d &= \frac{A\left(\frac{r}{n}\right)}{\left[\left(1+\frac{r}{n}\right)^{nt}-1\right]} \\ &= \frac{500000\left(\frac{0.08}{12}\right)}{\left[\left(1+\frac{0.08}{12}\right)^{12\cdot 30}-1\right]} \\ &\approx \$335.49 \end{aligned}$$

So, you would need to deposit \$335.49 each month to have \$500,000 in 30 years if your account earns 8% interest. □

A note about rounding.

If you are using the formulas and round during intermediate steps you will probably have some roundoff error. For this reason, we enter the whole expression into the calculator and do not show the intermediate steps.

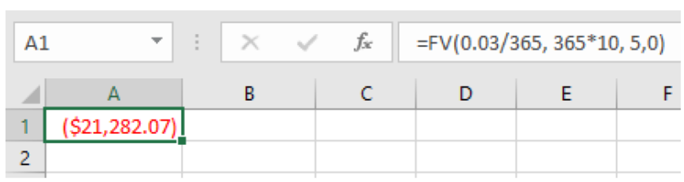
One of the challenges in this chapter is choosing the correct formula or spreadsheet function. Read this next example and see if you can determine which formula to use.

Example 2.3.4 A more conservative investment account pays 3% APR. If you deposit \$5 a day into this account, how much will you have after 10 years? What amount and percentage are from interest?

Solution. In this example we are given the regular deposit amount and we are looking for the future value. In a spreadsheet we use the =FV function and enter:

```
=FV(0.03/365, 365*10, 5,0)
```

which gives a result of 21,282.07.



To use a mathematical formula, we choose the one solved for A:

$d = \$5$ the daily deposit

$r = 0.03$ 3% annual rate

$n = 365$ since we're doing daily deposits, we'll compound daily

$t = 10$ we want the amount after 10 years

$$\begin{aligned} A &= \frac{5\left[\left(1+\frac{0.03}{365}\right)^{365\cdot 10}-1\right]}{\left(\frac{0.03}{365}\right)} \\ &\approx \$21,282.07 \end{aligned}$$

To find the amount of interest, we will calculate how much was deposited in the account. Since you put in \$5 a day for 10 years we get

$$\$5(365)(10) = \$18,250.$$

The interest earned is $\$21,282.07 - \$18,250 = \$3,032.07$

To find the percentage we divide by the total balance to get $\frac{\$3,032.07}{\$21,282.07} = 0.1424$ or 14.24%.

After 10 years, about 14.2% of the account is interest. □

2.3.3 Comparing Lump Sum and Regular Savings Payments

Now let's compare two scenarios to do some multistep problems and get a sense for the value of compounding over time.

Example 2.3.5 Scenario 1: Suppose you invest \$200 a month for 15 years into an account earning 10% APR compounded monthly. After 15 years, you leave the money, without making additional deposits, in the account for another 20 years. How much will you have in the end?

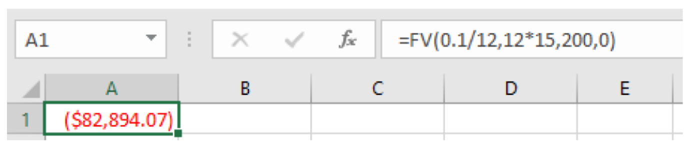
Scenario 2: Suppose instead you didn't invest anything for the first 15 years, then deposited \$200 a month for 20 years into an account earning 10% APR compounded monthly. How much will you have in the end?

Solution. Before we calculate the balance for both scenarios, which one do you think will have a higher balance at the end?

For scenario 1, there are two steps involved. The first part is the monthly payments for 15 years. To calculate this with a spreadsheet we enter

```
=FV(0.10/12, 12*15, 200,0)
```

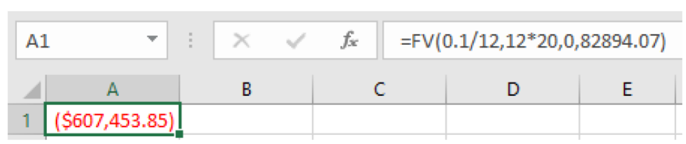
which gives \$82,894.07.



Now you will stop making payments and let the money sit and earn interest for 20 more years. With a spreadsheet we enter

```
=FV(0.10/12, 12*20, 0, 82894.07)
```

which gives \$607,453.85.



The process is similar with the formulas. For the first step we have:

$$\begin{aligned} A &= \frac{200\left[\left(1+\frac{0.10}{12}\right)^{12\cdot 15}-1\right]}{\left(\frac{0.10}{12}\right)} \\ &\approx \$82,894.07 \end{aligned}$$

And for the second step we use the compound interest formula from Section 2.2.

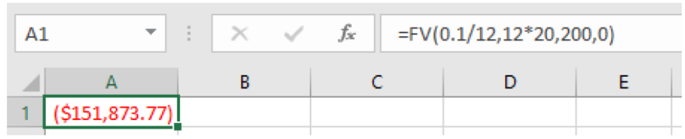
$$\begin{aligned} A &= \$82,894.07\left(1+\frac{0.10}{12}\right)^{12\cdot 20} \\ &\approx \$607,453.85 \end{aligned}$$

Now for Scenario 2: Since we are not investing anything for the first 15 years there is nothing to calculate.

This is a one-step problem. We will find the future value with the monthly payments of \$200 for 20 years. With a spreadsheet we enter

```
=FV(0.10/12, 12*20, 200,0)
```

which gives \$151,873.77.



To check that with the formula we have:

$$A = \frac{200\left[\left(1+\frac{0.10}{12}\right)^{12\cdot 20} - 1\right]}{\left(\frac{0.10}{12}\right)}$$
$$\approx \$151,873.77$$

Were you surprised by these numbers? You would put in less money in scenario 1 and end up with four times as much. The key to compounding interest is to start early. If you remember the graph of compound interest in Section 2.2, we can see that as time goes on, the balance increases exponentially. □

2.3.4 Deriving the Savings Plan Formula (Optional)

If you are interested in where the savings plan formula came from, we will explain it here. A savings plan with regular payments can be described recursively. Recall that basic compound interest follows from the relationship for each compound period.

$A = P\left(1+\frac{r}{n}\right)$

For a savings plan, we need to add a deposit, d, to the account with each compounding period:

$A = P\left(1+\frac{r}{n}\right) + d$

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.

Suppose we will deposit \$100 each month into an account paying 6% APR. We assume that the account is compounded with the same frequency as we make deposits unless stated otherwise.

In this example:

$r = 0.06$ 6% APR

$n = 12$ 12 compounds/deposits per year

$d = \$100$ our deposit per month

Writing out the recursive equation gives where A is exchanged with P_m where m is the number of compounding periods.

$P_m = \left(1+\frac{0.06}{12}\right) P_{m-1} + 100 = (1.005)P_{m-1} + 100$

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:

$P_m = 100(1.005)^{m-1} + 100(1.005)^{m-2} + ... + 100(1.005) + 100$

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. Last month's deposit would have earned only one month 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:

$1.005P_m = 1.005[100(1.005)^{m-1} + 100(1.005)^{m-2} + ... + 100(1.005) + 100]$

Distributing on the right side of the equation gives

$1.005P_m = 100(1.005)^m + 100(1.005)^{m-1} + ... + 100(1.005)^2 + 100(1.005)$

Now we'll line this up with like terms from our original equation, and subtract each side

$1.005P_m = 100(1.005)^m + 100(1.005)^{m-1} + ... + 100(1.005)$

$P_m = 100(1.005)^{m-1} + ... + 100(1.005) + 100$

Almost all the terms cancel on the right side when we subtract, leaving

$1.005P_m - P_m = 100(1.005)^m - 100$

Solving for P_m

$0.005P_m = 100((1.005)^m - 1)$

$P_m = \frac{100(1.005)^m - 100}{0.005}$

Replacing P_m with A (Future Value), m months with $12t$, where t is measured in years, gives

$A = \frac{100(1.005)^{12t} - 1}{0.005}$

Recall 0.005 was r/n and 100 was the deposit d. The value 12 as n, the number of deposits each year. Generalizing this result, we get the savings plan formula solved for A. The second formula uses algebra to rearrange the formula to be solved for d.

Savings Plan Formulas.

$$A = \frac{d\left[\left(1+\frac{r}{n}\right)^{nt} - 1\right]}{\left(\frac{r}{n}\right)} \text{ or } d = \frac{A\left(\frac{r}{n}\right)}{\left[\left(1+\frac{r}{n}\right)^{nt} - 1\right]}$$

A is the balance in the account after n years (future value)

d is the regular deposit (or payment amount each month, quarter, year, etc.)

r is the annual interest rate in decimal form

n is the number of compounding periods in one year

t is the number of years

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

2.3.5 Exercises

1. You set up a savings plan for retirement in 35 years. You will deposit \$250 each month for 35 years. The account will earn an average of 6.5% APR compounded monthly.

a. How much will you have in your retirement plan in 35 years?

b. How much interest did you earn.

c. What percent of the balance is interest?

2. You set up a savings plan for retirement in 40 years. You will deposit \$75 each week for 40 years.The account will earn an average of 8.5% APR compounded weekly.

a. How much will you have in your retirement plan in 40 years?

b. How much interest did you earn.

c. What percent of the balance is interest?

3. You set up a savings plan for retirement in 30 years. You will deposit \$750 each quarter for 30 years. The account will earn an average of 7.75% APR compounded quarterly.

a. How much will you have in your retirement plan in 30 years?

b. How much interest did you earn?

c. What percent of the balance is interest?

4. You set up a savings plan for retirement in 25 years. You will deposit \$20 per day for 25 years. The account will earn an average of 2.35% APR compounded daily.

a. How much will you have in your retirement plan in 25 years?

b. How much interest did you earn?

c. What percent of the final balance is interest?

5. Suppose you invest \$130 a month for 5 years into an account earning 9% APR compounded monthly. After 5 years, you leave the money, without making additional deposits, in the account for another 25 years.

a. How much will you have in the end?

b. How much interest did you earn?

c. What percent of balance is interest?

6. Suppose you invest \$200 per month for 10 years into an account earning 5% APR compounded monthly. You then leave the money, without making additional deposits, in the account for another 20 years.

a. How much will you have after the first 10 years?

b. How much will you have after the additional 20 years?

c. How much total interest did you earn?

d. What percent of the final balance is interest?

7. Suppose you have 30 months in which to save \$3,500 for a cruise for your family. If you can earn an APR of 3.8%, compounded monthly, how much should you deposit each month?

8. You wish to have \$3,000 in 2 years to buy a fancy new stereo system. How much should you deposit each quarter into an account paying 6.5% APR compounded quarterly?

9. Jamie has determined they need to have \$450,000 for retirement in 30 years. Their account earns 6% APR. How much would Jamie need to deposit in the account each month?

10. Lashonda already knows that she wants \$500,000 when she retires. If she sets up a saving plan for 40 years in an account paying 10% APR, compounded quarterly, how much should she deposit each quarter?

11. Jose' inherits \$55,000 and decides to put it in the bank for the next 25 years to save for his retirement. He will earn an average of 5.6% APR compounded monthly for the next 25 years. His partner deposits \$375 a month in a separate savings plan that earns 5.6% APR compounded monthly for the next 25 years.

a. How much will each have at the end of 25 years?

b. How much interest did each person earn?

c. What percent of balance is interest for each person?

12. Akiko inherits \$45,000 and decides to put it in the bank for the next 30 years to save for her retirement. She will earn an average of 7.8% APR compounded monthly for the next 30 years. Her spouse deposits \$200 a month in a separate savings plan that earns 7.8% APR compounded monthly for the next 30 years.

a. How much will each have at the end of 30 years?

b. How much interest did each person earn?

c. What percent of balance is interest for each person?

13. Sylvin makes an initial deposit of \$1000 into a savings account and then adds \$100 each month for 10 years into an account pays 4.5% APR compounded monthly.

a. What will be her final balance?

b. How much interest did she earn?

c. What percent of the final balance is interest?

14. Elena makes an initial deposit of \$5000 into a savings account and then adds \$1000 each year for 20 years into an account pays 2.35% APR compounded annually.

a. What will be her final balance?

b. How much interest did she earn?

c. What percent of her final balance is interest?

15. Vanessa just turned 40 years old. Her plan is to save \$100 per month until retirement at age 65. Suppose she deposits that \$100 each month into a savings account that earns 4% APR compounded monthly.

a. What will her balance be when she turns 65 years old?

b. If she started saving when she turned 25 years old instead, what would her balance be?

16. Chris wants to start saving money for retirement. Suppose he deposits \$1000 every year into a savings account that pays 5% APR compounded annually.

a. How much will Chris have saved in 20 years?

b. How much will Chris have saved in 40 years?

17. Fareshta and Ahmad want to save to help send their child to college. Their plan is to put aside \$50 every week. Suppose they deposit that money into an account that pays 3.5% APR compounded weekly.

a. How much money will be in the account in 18 years? (assume 52 weeks in a year)

b. What minimum initial lump sum deposit would they need to make today to have the same balance in 18 years if they weren't putting aside the \$50 per week?

18. Elisa decides to cancel her cable TV and to deposit the \$100 she will save each month into an account that pays 4.5% APR compounded monthly.

a. How much will she have in the account in 10 years?

b. What minimum initial lump sum deposit would she need to make today to have the same balance in 10 years without saving the $100 per month?

2.4 Loan Payments

Objectives: Section 2.4 Loan Payments

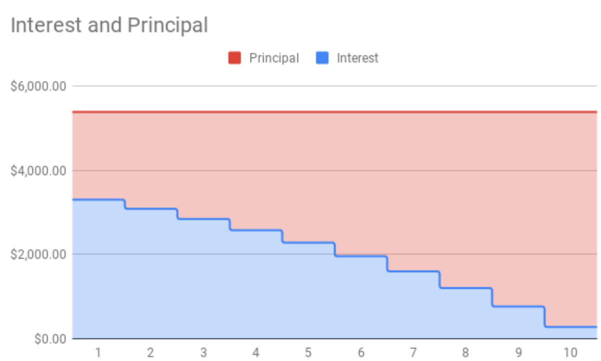
Students will be able to:

- Use a spreadsheet and/or formula to calculate the payment amount for student loans, car loans, paying off credit cards and mortgage loans
- Calculate the total paid over the life of a loan, amount of interest paid, and the percentage of the total amount paid in interest
- Determine when to use each formula in the financial math chapter

In the last section, you learned about savings plans. In this section, you will learn about conventional loans (also called amortized loans or installment loans). Examples include student loans, car 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. In this section we will also briefly cover credit cards. There is a more thorough chapter on credit cards available as an optional chapter.

2.4.1 Installment Loans

Installment loans are also called ***amortized loans***. They are designed so that the payment amount remains the same over the life (term) of the loan. In order for that amount to remain constant, the amount of money going towards paying the ***principal*** (amount owed) and towards the ***interest*** will vary. At the beginning of the loan, the amount owed is the largest. So, at the beginning the amount of interest charged will also be the largest. As the loan gets repaid the amount owed is reduced and therefore the interest is reduced. Therefore, at the beginning of an installment loan payments go mostly towards the interest, towards the end the payments are going mostly towards the principal.



This is a graph of the interest and principal paid on a loan of \$34,000 at 10% APR. Each year \$5391.72 was paid in total. In the first year, \$3,3305.92 was applied to the interest and \$2085.80 was applied to the principal. In the last year of the loan, only \$281.04 was applied to the interest and \$5,110.68 was applied to the principal. Over 5 years, a total of \$53,917.2 was paid for a loan of \$34,000.

2.4.2 Loan Formulas

In a savings plan, you start with nothing, put money into an account once or on a regular basis, and have a larger balance at the end. Loans work in reverse. You start with a balance owed, make payments and the future value is zero when the loan is paid off.

We will continue to use the same spreadsheet formulas. The ones that are most useful for loans are `=PV` and `=PMT`. We will look at how the inputs change for a loan.

Spreadsheet Formulas.

```
=PV(rate per period, number of periods, payment amount, future value)

=PMT(rate per period, number of periods, present value, future value)
```

rate per period is the interest rate per compounding period, r/n

number of periods is the total number of periods, $n * t$

payment amount is the amount of regular payments, d

present value is the amount deposited or principal, P

future value is the amount you want in the future, 0 for a loan

These two formulas correspond to the formulas below. The formula for loans is derived in a similar way that we did for savings plans, but notice they have negative exponents. The details are omitted here.

Loan Formulas.

$$P = \frac{d\left(1-\left(1+\frac{r}{n}\right)^{-nt}\right)}{\left(\frac{r}{n}\right)} \text{ or } d = \frac{P\left(\frac{r}{n}\right)}{\left(1-\left(1+\frac{r}{n}\right)^{-nt}\right)}$$

P 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

n is the number of compounding periods in one year

t 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

Example 2.4.1 Teresa wants to buy a car that costs $15,000. She has $3,000 saved for the car and plans to finance the rest. She found a 3-year loan at 2.75% APR and a 5-year loan at 4% APR. How much will her monthly car payment be for each loan and how do these loans compare to each other.

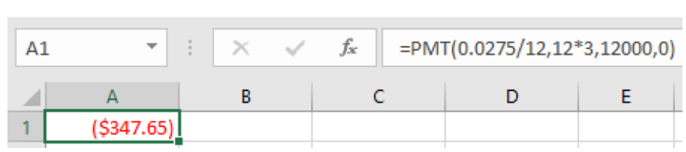
To use a spreadsheet, we use the `=PMT` formula. For a loan, the loan amount is the present value and the future value is 0, indicating that the loan will be paid off. Teresa is making a down payment, so we also need to subtract that from the cost of the car to find the loan amount:

$\$15,000\breve{\ }\$3,000 = \$12,000$

Her loan amount is $12,000. For the 3-year loan at 2.75% APR, we enter:

```
=PMT(0.0275/12, 12*3, 12000, 0)
```

and get a result of $347.65.



For the formula, we use the one solved for d:

$r = .0275$ for 2.75% annual rate

$n = 12$ monthly payments

$t = 3$ for 3 years

$P = 12000$ Since she can pay $3,000 of the $15,000

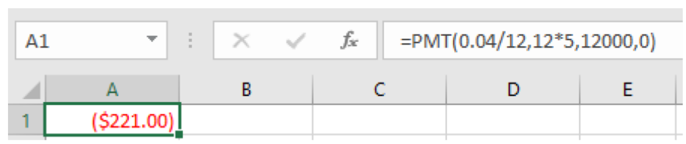
$$
\begin{aligned}
d &= \frac{P\left(\frac{r}{n}\right)}{\left(1-\left(1+\frac{r}{n}\right)^{-nt}\right)} \\
&= \frac{12000\left(\frac{0.0275}{12}\right)}{\left(1-\left(1+\frac{0.0275}{12}\right)^{-12\cdot 3}\right)} \\
&\approx \$347.65
\end{aligned}
$$

Teresa's car payment would be $347.65.

Now for the 5-year loan at 4% APR, we enter:

```
=PMT(0.04/12, 12*5, 12000, 0)
```

and we get $221.00.



To use the formula, we have:

$r = .04$ for 4% annual rate

$n = 12$ monthly payments

$t = 5$ for 5 years

$P = 12000$ the loan amount

$$
\begin{aligned}
d &= \frac{P\left(\frac{r}{n}\right)}{\left(1-\left(1+\frac{r}{n}\right)^{-nt}\right)} \\
&= \frac{12000\left(\frac{0.04}{12}\right)}{\left(1-\left(1+\frac{0.04}{12}\right)^{-12\cdot 3}\right)} \\
&\approx \$221.00
\end{aligned}
$$

Now let's compare the loans by finding out how much Teresa would pay in interest for each loan.

For the 3-year loan at 2.75% APR, her payments would total:
$\$347.65(12)(3) = \$12,515.40$. Her interest would be $515.40.

For the 5-year loan at 4% APR, her payments would total:
$\$221.00(12)(5) = \$13,260.00$. Her interest would be $1,260.00.

There are two main differences between these two loans: the monthly payments and the total paid over the life of the loans. The first loan has a higher monthly payment by $126.65 per month. However, she would pay $744.60 less in interest. □

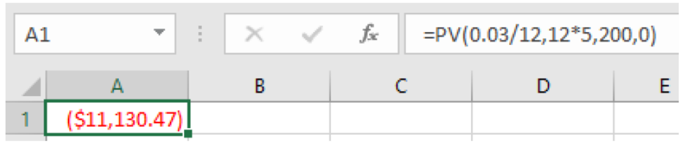
In addition to loan payments, we can calculate the amount of loan we can afford given a monthly payment. Let's look at that in the next example.

Example 2.4.2 You can afford \$200 per month as a car payment. If you can get an auto loan at 3% APR 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?

Solution. To use a spreadsheet for this problem, we use the `=PV` formula because we want to know the present value, which is the value of the loan right now. We enter

```
=PV(0.03/12, 12*5, 200,0)
```

which gives \$11,130.47.



To use a formula, we are looking for P, the starting amount of the loan.

$d = \$200$ the monthly loan payment

$r = 0.03$ 3% annual rate

$n = 12$ since we're doing monthly payments, we'll compound monthly

$t = 5$ since we're making monthly payments for 5 years

$$
\begin{aligned}
P &= \frac{d\left(1-\left(1+\frac{r}{n}\right)^{-nt}\right)}{\frac{r}{n}} \\
&= \frac{200\left(1-\left(1+\frac{0.03}{12}\right)^{-12\cdot 5}\right)}{\frac{0.03}{12}} \\
&= \frac{200(1-(1.0025)^{-60})}{0.0025} \\
&\approx \$11,130.47
\end{aligned}
$$

You can afford a maximum loan of \$11,130.47. If you have a down payment you can add that to get the value of the car you can buy. If there are any closing costs for the loan you also need to take that into consideration.

To find the amount of interest you will pay for this loan, calculate the total of all your payments.

$\$200(5)(12) = \$12,000$

Then take the difference between the total payments and the loan amount.

$\$12,000 - \$11,130.47 = \$869.53$

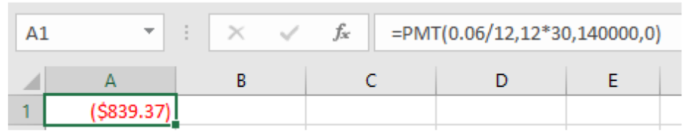
In this case, you would be paying \$869.53 in interest. □

So far, we have looked at car loans. Student loans and home mortgages are calculated in the same way. Here is an example of a mortgage payment.

Example 2.4.3 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? What percentage of your total payments will go towards interest?

Solution. To use a spreadsheet for this problem, we use the `=PMT` formula because we want to know the payment amount. The amount of the loan is the present value and to pay off the loan the future value is 0. We enter

```
=PMT(0.06/12, 12*30, 140000,0)
```



which gives us \$839.37.

To use the formula, we have:

$r = 0.06$ 6% annual rate

$n = 12$ since we're paying monthly

$t = 30$ 30 years

$P = \$140,000$ the starting loan amount

In this case, we're going to use the equation that is solved for d.

$$\begin{aligned} d &= \frac{P\left(\frac{r}{n}\right)}{\left(1-\left(1+\frac{r}{n}\right)^{-nt}\right)} \\ &= \frac{140000\left(\frac{0.06}{12}\right)}{\left(1-\left(1+\frac{0.06}{12}\right)^{-12\cdot 30}\right)} \\ &= \frac{700}{(1-(1.005)^{-360})} \\ &\approx \$839.37 \end{aligned}$$

You would make payments of \$839.37 per month for 30 years. To find out what percentage of the total will go towards interest, we need to total up all of the payments.

$$\$839.37(30)(12) = \$302,173.20$$

Then take the difference between the total payments and the loan amount.

$$\$302,173.20 - \$140,000 = \$162,173.20$$

In this case, you would be paying \$162,173.20 in interest over the life of the loan. To find the percentage, we divide the interest by the total amount paid.

$$\frac{\$162,173.20}{\$302,173.20} \approx 0.5367$$

About 53.7% of the total is being paid towards interest. □

2.4.3 Remaining Loan 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 will have to pay from the sale.

To determine the remaining loan balance after some number of years, we first need to calculate the payment amount, if we don't already know it. 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 2.4.4 If a 30-year mortgage at an interest rate of 6% APR has payments of $1,000 a month, what will the loan balance be in 5 years?

Solution. To determine this, we need to think backwards. We are looking for the amount of the loan that can be paid off by $1,000 per month in the remaining 25 years. In other words, we're looking for P when:

$d = \$1,000$ for the monthly loan payment

$r = 0.06$ for 6% annual rate

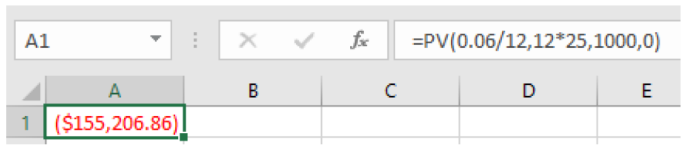
$n = 12$ since we're doing monthly payments, we'll compound monthly

$t = 25$ since we'd be making monthly payments for 25 more years

To use a spreadsheet for this problem, we use the =PV formula because we want to know what the present value would be at the time you want to sell in 5 years. We enter:

```
=PV(0.06/12, 12*25, 1000,0)
```

which gives us $155,206.86.



To check this with the formula we have:

$$
\begin{aligned}
P &= \frac{1000\left(1-\left(1+\frac{0.06}{12}\right)^{-12\cdot 25}\right)}{\frac{0.06}{12}} \\
&= \frac{1000(1-(1.005)^{-300})}{0.005} \\
&\approx \$155,206.86
\end{aligned}
$$

The loan balance with 25 years remaining on the loan will be $155,206.86 □

Sometimes answering remaining balance questions requires two steps, both of which we have done in this section:

1 Calculate the monthly payment on the loan

2 Calculate the remaining loan balance based on the *remaining time* on the loan

2.4.4 Credit Cards

Credit cards are useful to many people. They can be used to build a credit score, as a short-term loan and as an alternative to physical cash. It is highly advisable to fully pay off a credit card each month because the interest rates are much higher than the conventional installment loans we discussed earlier. There can also be many expensive fees applied when carrying a balance. Because of these higher rates, it can be very easy to get into a lot of debt quickly.

Credit cards are a common source of loan money, and therefore a common source of debt which must be repaid. There are much fewer and lower barriers to obtaining and using a credit card than there are to obtaining, for example, a mortgage. As such, there are some drawbacks both in terms of typically higher interest rates, as well as some structural differences.

Example 2.4.5 Bilqis gets a credit card, with 17.99% APR compounding daily. She uses the card to buy a \$900 plane ticket. She does not make any payments during the interest-free grace period. Her first payment is due 35 days after the grace period ended. Following the fine print in the credit card agreement, the minimum payment is calculated to be 1% of the outstanding balance after applying the APR; or \$25, whichever is larger. What is her outstanding balance? What is her minimum payment? If she pays the minimum payment what will her new balance be?

Solution. First we will see what she owes with the 17.99% APR using the compound interest formula for 35 days:

$$900\left(1+\frac{0.1799}{365}\right)^{35} \approx 915.66$$

The spreadsheet formula for this computation is

```
=FV(0.1799/365,35,0,900)
```

With interest, Bilqis now owes \$915.66.

Now we will find one percent of the new balance which is

$$\$915.66(0.01) \approx \$9.16.$$

\$25 is larger than \$9.16, so her first payment must be at least \$25. If she makes the minimum payment we can subtract to find her new balance:

$$\$915.66 - \$25 = \$890.66.$$

Her new balance will be \$890.66. □

2.4.5 Summary of Spreadsheet Formulas

Here are all the spreadsheet formulas from this chapter so far together so you can see the similarities and differences.

Spreadsheet Formulas.

```
=principal+principal*rate*time

=FV(rate per period, number of periods, payment amount, present value)

=principal*EXP(yearly rate*years)

=PV(rate per period, number of periods, payment amount, future value)

=PMT(rate per period, number of periods, present value, future value)

=EFFECT(stated rate, number of compounding periods per year)
```

rate per period is the interest rate per compounding period, r/n

number of periods is the total number of periods, $n * t$

payment amount is the amount of regular payments, d

present value is the amount deposited or principal, P

future value is the amount you want in the future, 0 *for a loan*

2.4.6 When to use the formulas: What is the question asking?

- Find a payment: `=PMT`
- Find the effective rate or compare accounts: `=EFFECT`
- How much do you need to deposit now, what loan amount can you afford, or remaining loan balance: `=PV`
- What will the account balance be in the future?
  - Simple interest: `=principal+principal*rate*time`
  - Compound interest (except continuous): `=FV`
  - Continuously compounded interest: `principal*EXP(rate*years)`

2.4.7 Summary of Mathematical Formulas

Mathematical Formulas.

Simple Interest

$$I = Prt \text{ or } A = P + Prt$$

Compound Interest

$$A = P\left(1 + \frac{r}{n}\right)^{nt} \text{ or } P = \frac{A}{\left(1 + \frac{r}{n}\right)^{nt}}$$

Continuously Compounded

$$A = Pe^{rt} \text{ or } P = \frac{A}{e^{rt}}$$

Savings Plans

$$A = \frac{d\left[\left(1 + \frac{r}{n}\right)^{nt} - 1\right]}{\left(\frac{r}{n}\right)} \text{ or } d = \frac{A\left(\frac{r}{n}\right)}{\left[\left(1 + \frac{r}{n}\right)^{nt} - 1\right]}$$

Loans

$$P = \frac{d\left(1 - \left(1 + \frac{r}{n}\right)^{-nt}\right)}{\frac{r}{n}} \text{ or } d = \frac{P\left(\frac{r}{n}\right)}{\left(1 - \left(1 + \frac{r}{n}\right)^{-nt}\right)}$$

P is the principal, starting amount, or present value

d is your loan payment (your monthly payment, annual payment, etc.)

r is the annual intrest rate in decimal form

n is the number of compounding periods in one year

t is the length of the loan, in years

A is the end amunt or future value

If the compounding frequency is not always explicitly given, it is determined by how often you make payments

2.4.8 When to use the formulas: What is the question asking?

- Find a payment
  - Savings payment: savings plan equation (positive exponent) solved for d
  - Loan payment: loan equation (negative exponent) solved for d
- How much do you need to deposit now?
  - Compound interest (except continuous): compound interest formulas solved for P
  - Continuously compounded: the formula with e solved for P
- What loan amount can you afford, or remaining loan balance: loan formula solved for P
- What will the account balance be in the future?
  - One-time deposit:
    - Simple interest: simple interest formula
    - Compound interest (except continuous): compound interest formula solved for A
    - Continuously compounded interest: the formula with e in it, solved for A
  - Regular payments: Savings plan formula solved for A

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 to determine what approach will best allow you to solve the problem. After practicing with the exercises from this section, you can test yourself on which formula to use with exercises 14-21.

2.4.9 Exercises

1. You can afford a \$700 per month mortgage payment. You've found a 30-year loan at 5.5% APR.

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?

2. Marie can afford a \$250 per month car payment. She's found a 5-year loan at 7% APR.

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?

3. You want to buy a $25,000 car. The company is offering an interest rate of 2% APR for 48 months (4 years). What will your monthly payments be?

4. You decide to finance a $12,000 car at 3% APR compounded monthly for 4 years. What will your monthly payments be? How much interest will you pay over the life of the loan?

5. 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%?

6. Lynn bought a $300,000 house, paying 10% down, and financing the rest at 6.5% APR for 30 years.

a. Find her monthly payments.

b. How much interest will she pay over the life of the loan?

c. What percentage of her total payment was interest?

7. Emile bought a car for $24,000 three years ago. The loan had a 5-year term at 3% APR. How much does he still owe on the car?

8. A friend bought a house 15 years ago, taking out a $120,000 mortgage at 6% APR for 30 years. How much does she still owe on the mortgage?

Use a spreadsheet (not formulas) to answer problems 9 through 13 below.

9. Imagine you have $20,000 saved as a down payment on a house. You wish to take out a fixed-rate 30-year mortgage loan at 4% APR (remember that mortgage rates usually assume monthly compounding). If the maximum mortgage payment you can afford is $950 per month, then what is the maximum house price that you can afford?

10. Suppose another mortgage lender offers you a fixed-rate 15-year mortgage at 2.95% APR. You have $20,000 saved as a down payment, and you can afford a maximum mortgage payment of $950 per month. You are interested in a certain house for sale, with firm selling price of $200,000.

a. Find the monthly payment for this house. Can you afford it, under the terms of this lender?

b. (Challenge): Suppose this same lender offers to increase the APR by only 0.05%, for each additional year added to the loan period beyond 15 years (so that a 16-year loan would have 3.00% APR, and a 17-year loan would have 3.05% APR, and so on), up to a maximum loan period of 25 years. Given these terms, does any combination of APR and loan period exist that would let you afford the house? If so, state the *minimum* number of additional years needed, the total resulting loan period, the resulting APR, and the resulting monthly payment.

11. An annuity firm pays 5% APR compounded yearly, and offers an investor the following: Deposit $100,000 with them today, and then starting one year from today, you will receive ten equal annual payments, with zero balance remaining afterward in the account. If an investor accepts their offer, then find the following:

a. What will be the payment amount for each of the ten equal payments?

b. After ten years, how much interest will the investor have received, and what percentage of the total payment sum will represent interest?

12. For twelve full years, and into an account that pays 3.5% APR compounded quarterly: Yanhong will either pay $1500 at the end of each calendar quarter, or, deposit a single lump sum that will give the same future value amount.

a. If Yanhong chooses the single lump sum option, then how much will Yanhong need to deposit?

b. If Yanhong needs to have earned \$100,000 in this account at the end of the twelve years, then the quarterly deposit amount will need to be increased. What would the new quarterly deposit amount need to be?

c. (Challenge): If Yanhong will make quarterly deposits into this account for the twelve years, but also has \$8,000 to additionally deposit into this account right away: What would the new quarterly deposit amount need to be, so that the total balance after twelve years is \$100,000?

13. Assume you take out a 30-year mortgage loan for \$250,000 at a fixed 4.5% APR.

a. What will be the amount of your monthly payment?

b. After the first ten years of payments, how much will remain on your loan balance?

c. After the first twenty years of payments, how much will remain on your loan balance?

d. Notice that the amount of the loan balance reduction during the second ten years, was very considerably bigger than the amount of the loan balance reduction during the first ten years. Why does the loan balance decrease at a faster and faster pace, the longer that the loan has been in repayment?

For each of the following scenarios, determine which formula from sections 2.2-2.4 to use and solve the problem.

14. Keisha received an inheritance of \$20,000 and invested it at 6.9% APR, compounded continuously. How much will she have for college in 8 years?

15. Paul wants to buy a new car. Rather than take out a loan, he decides to save \$200 a month in an account earning 3.5% APR compounded monthly. How much will he have saved up after 3 years?

16. Sol is managing investments for a non-profit company. They want to invest some money in an account earning 5% APR compounded annually with the goal to have \$30,000 in the account in 6 years. How much should Sol deposit into the account?

17. Miao is going to finance new office equipment at 2.8% APR over a 4-year term. If she can afford monthly payments of \$100, how much equipment can she buy?

18. How much would you need to save every month in an account earning 4.1% APR to have \$5,000 saved up in two years.

19. Terry and Jess are buying a house for \$405,000 and they can afford to put 10% down. Their interest rate is 4.3% APR for 30 years. What will their monthly mortgage payment be?

20. You loan your sister \$500 for two years and she agrees to pay you back with 3% simple interest per year. How much will she pay you back

21. Zahid starts saving \$150 per month in an account that pays 4.8% APR compounded monthly. If he continues for 20 years, how much will he have? If he waited 10 years instead and put in \$300 per month for 10 years with the same interest, how much would he have?

2.5 Income Taxes

Objectives: Income Taxes 2.5

Students will be able to:

- Calculate gross income and adjusted gross income (AGI)
- Determine the standard deduction according to filing status
- Determine whether to use the standard or itemized deductions and calculate taxable income
- Calculate income tax from tables
- Compare taxes owed to withholdings to determine whether a refund is due or a payment is required

2.5.1 A Very Brief History of Taxes

Although Benjamin Franklin famously claimed in 1789 that "in this world nothing can be said to be certain, except death and taxes", it wasn't until 1913 that the 16th amendment was ratified, and income tax was legalized. Prior to this, taxes were primarily collected through tariffs on imported goods, poll taxes, and property taxes. A ***poll tax***, also referred to as a head tax, was a fixed amount every liable individual had to pay. Payment of the poll tax was often required before a person could register to vote or be issued a hunting or fishing license.

Tax policy in the United States is a politically divisive issue. In general, Democrats seek to lower taxes for low and middle income earners and raise taxes for the wealthy, while Republicans support a variety of tax cuts and limiting the amount wealthy earners are taxed. There is even debate as to the current length of the tax code! Some claim that the code is over 70,000 pages while others insist it is just over 2,000. Nevertheless, there is at least one thing everyone can agree with - you don't want to be on the wrong side of the Internal Revenue Service (IRS)!

2.5.2 Types of Income Tax

Income tax is a tax that is levied on earned income or profit. Income taxes are collected by the federal government, states, and even some municipalities. Income taxes are an important source of funding, and are used to finance social programs, maintain and expand infrastructure, and provide foreign aid, among other things.

2.5.3 Federal Income Tax

Anyone who earns income over a certain amount (approximately $10,000 for an individual) must file a federal tax return and have taxes collected on their behalf. The amount of federal tax owed is determined by your ***filing status*** and ***taxable income***.

2.5.4 State Income Tax

Forty-one states and the District of Columbia collect taxes on income from wages and investment. Seven states - Washington, Nevada, South Dakota, Wyoming, Alaska, Texas, and Florida – do not collect tax on wage or investment income, and two states – New Hampshire and Tennessee – only collect tax on investment income. The amount of tax collected varies by state, with each state averaging between $500 and $3000 per person.

2.5.5 Municipal Income Tax

Some municipalities (urban districts) also collect income taxes. For example, business owners in Multnomah County pay a municipal tax on their business income, and workers in the Tri-met district (area served by Tri-met) have their wages taxed at 0.7537%.

2.5.6 Calculating Federal Income Tax

Not all income is taxed, and not all income that gets taxed is taxed at the same rate. The general process of calculating the amount of federal income tax you owe is outlined in the chart below.

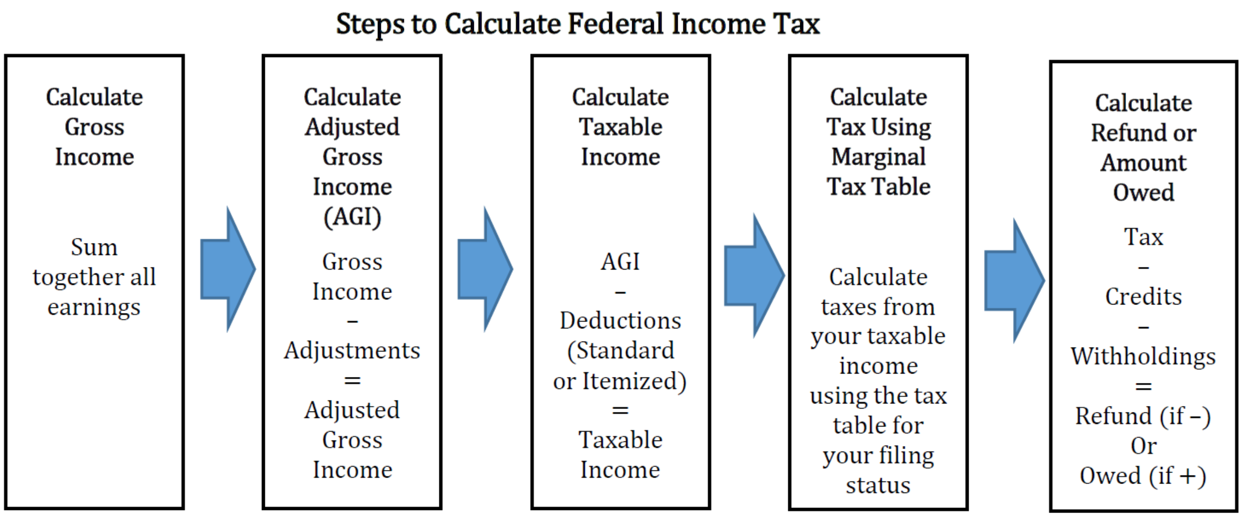


Figure 2.5.1 Steps to Calculate Federal Income Tax

2.5.7 Gross Income

Gross income includes all wages, tips, earned interest, dividends, rents and royalties, alimony, property gains, income tax refunds, etc. Keep in mind that ALL income means ALL income. Even income earned from a crime must be reported!

2.5.8 Adjusted Gross Income

Adjustments are eligible expenses used to reduce your gross income. Adjustments are tax-exempt and thus reduce the amount owed in taxes. Eligible expenses include contributions to tax deferred savings plans (401k, Individual Retirement Plan (IRA)), school tuition, student loan interest, moving expenses, business expenses, flexible spending accounts and health savings account contributions.

2.5.9 Taxable Income

Taxable income is determined by subtracting your ***deductions*** from your adjusted gross income. You may choose to take *either* a ***standard deduction***, which is determined by your filing status, or you may choose to ***itemize*** your deductions. Itemized deductions may include state and local income taxes, property taxes, medical expenses, mortgage interest, and charitable donations.

Prior to the 2018 tax year, you could also claim ***personal exemptions***. Exemptions of $4,050 for each member of the household were subtracted from the adjusted gross income along with either the standard or itemized deduction to determine taxable income. Starting in 2018, however, the standard deduction was doubled, and personal exemptions were eliminated.

Let's look at an example of how to calculate gross income and adjusted gross income (AGI).

Example 2.5.2 Sasha received $45,000 in wages and earned $1,300 in interest from his savings plan. He paid $1,400 in student loan interest, and put $4,000 into his Individual Retirement Account (IRA). Determine Sasha's adjusted gross income.

To calculate Sasha's adjusted gross income, we need to first determine his gross income. His gross income includes his wages and earned interest:

Gross Income $= \$45,000 + \$1,300 = \$46,300$

To find his adjusted gross income, we need to subtract eligible adjustments from his gross income. His eligible adjustments include the interest paid on his student loans, and his contributions to his IRA:

Adjusted Gross Income $= \$46,300 - \$1,400 - \$4,000 = \$40,900$ □

2.5.10 Filing Status

Your ***filing status*** is determined by your family situation. Are you married, widowed, divorced, caring for a family member? It is possible to fall into more than one category, but you may choose the one that is most beneficial for you.

- ***Single:*** If you are unmarried, legally separated from your spouse, or divorced on the last day of the year.
- ***Married Filing Jointly:*** If you are married and both you and your spouse agree to file a joint return. (On a joint return, you report your combined income and deduct your combined allowable expenses.)
- ***Married Filing Separately:*** If you are married and you want to be responsible only for your own tax, or if this status results in less tax than a joint return.
- ***Head of Household (with qualifying person):*** If you are 1) unmarried or considered unmarried on the last day of the year, 2) paid more than half the cost of keeping up a home for the year, and 3) have a qualifying dependent living with you for more than half the year (except temporary absences, such as school).
- ***Qualifying Widow/Widower (with dependent child):*** This status is available for the two years following the death of your spouse.

Your filing status determines your standard deduction and tax liability. The standard deduction for each filing status for the tax year 2018 is given in the table below.

Filing Status	Standard Deduction for Tax Year 2018
Single	$12000
Married Filing Jointly	$24000
Married Filing Separately	$12000
Head of Household	$18000
Qualifying Widow	$24000

Here is an example where we calculate adjusted gross income and taxable income.

Example 2.5.3 Maria earned wages of $32,400 from her job as a server. She also earned $8,300 in tips, and received $95 in interest from a savings account. In trying to save for retirement, Maria has put $3,500 into a 401K tax deferred savings plan. She is unmarried and lives with her six year old daughter. Determine Maria's gross income, adjusted gross income, and taxable income.

Solution. Maria's gross income includes her wages, tips, and earned interest.

Gross Income = \$32,400 + \$8,300 + \$95 = \$40,795

Maria can use her 401k contribution as an adjustment.

Adjusted Gross Income = \$40,795 - \$3,500 = \$37,295

Since Maria is unmarried and is the primary caregiver for her daughter, she can file as the head of household and take the standard deduction of \$18,000.

Taxable Income = \$37,295 - \$18,000 = \$19,295 □

2.5.11 Tax Tables

The United States has a ***progressive*** federal income tax system which means the more income you have, the more you generally pay in taxes. There are ***marginal tax rates*** that go up according to our income, but we don't pay the same rate on all of our income. The ranges of income are called ***tax brackets*** and the buckets in the illustration below can help us visualize the brackets.

The cutoff values in the figure are for a married couple filing jointly, but the percentages are the same for everyone as you'll see in the table on the next page. We pay 10% of any taxable income in the first bucket. If the first bucket is full, we move to the second bucket and we pay 12% of that income. This couple is in the 22% tax bracket because the 3rd bucket is partially filled. They will pay 22% of the income in that bucket, but they are not paying 22% of their entire income.

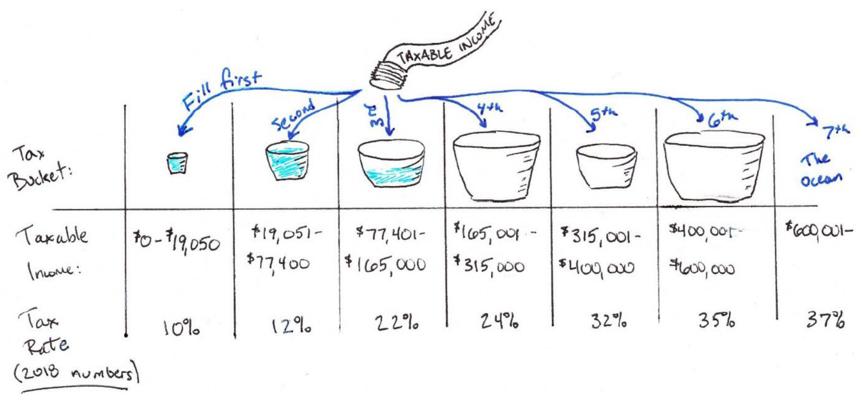


Figure 2.5.4 Tax Buckets by John Chesbrough, licensed under CC-BY-ND-NC 4.0.

Here is a table with cutoff values for each filing status. Everyone pays a tax of 10% on any income in the first tax bracket. Then we all pay 12% of the next bracket of income, 22% of the next bracket, and so on. This keeps going up to the 37% tax bracket.

2018 Tax Year				
Marginal Tax Rate On Taxable Income	Filing Single	Filing as Head of Household	Married filing Jointly	Married filing Separately
10%	First \$9,525	First \$13,600	First \$19,050	First \$9,525
12%	\$9,525 – \$38,700	\$13,600 – \$51,800	\$19,050 – \$77,400	\$9,525 – \$38,700
22%	\$38,700 – \$82,500	\$51,800 – \$82,500	\$77,400 – \$165,000	\$38,700 – \$82,500
24%	\$82,500 – \$157,500	\$82,500 – \$157,500	\$165,000 – \$315,000	\$82,500 – \$157,500
32%	\$157,500 – \$200,000	\$157,500 – \$200,000	\$315,000 – \$400,000	\$157,500 – \$200,000
35%	\$200,000 – \$500,000	\$200,000 – \$500,000	\$400,000 – \$600,000	\$200,000 – \$300,000
37%	Over \$500,000	Over \$500,000	Over \$600,000	Over \$300,000

Let's see how to use this table to calculate taxes in an example.

Example 2.5.5 If Avery is filing single and has \$55,100 in taxable income, calculate their tax.

Solution. We begin with the lowest tax bracket and take 10% of \$9,525. Since their income is higher than that, we add 12% of the next amount, found by subtracting the values in that bracket. Avery's taxable income is in the 22% tax bracket, so we find the amount over \$38,700 by subtracting. Here is the full calculation:

$$\begin{aligned} 0.10(\$9,525) + 0.12(\$38,700 - \$9,525) + 0.22(\$55,100 - \$38,700) &= 0.10(\$9,525) + 0.12(\$29,175) + 0.22(\$16,400) \\ &= \$952.50 + \$3,501 + \$3,608 \\ &= \$4,435.50 + \$3,608 \\ &= \$8,043.50 \end{aligned}$$

Avery would owe \$8,043.50 in taxes. Note that their overall tax rate is somewhere between 10% and 22%. We can calculate the actual rate by dividing their tax owed by their taxable income:

$\frac{\$8,043.50}{\$55,100} = 0.1459$ or about 15% □

We could calculate all taxes this way, but you might notice that the first few terms will be the same if the buckets are full. For this reason, we can simplify these tax tables. The tax tables below give the value for all of the lower tax buckets that are full. There is a tax table for each filing status and the cutoffs are regularly adjusted for inflation, so they usually vary from year to year.

2018 Federal Income Tax Tables

Single (2018)	
If taxable income is:	The tax is:
Not over \$9,525	10% of the taxable income
Over \$9,525 but not over \$38,700	\$952.50 plus 12% of the excess over \$9,525
Over \$38,700 but not over \$82,500	\$4,453.50 plus 22% of the excess over \$38,700
Over \$82,500 but not over \$157,500	\$14,089.50 plus 24% of the excess over \$82,500
Over \$157,500 but not over \$200,000	\$32,089.50 plus 32% of the excess over \$157,500
Over \$200,000 but not over \$500,000	\$45,689.50 plus 35% of the excess over \$200,000
Over \$500,000	\$150,689.50 plus 37% of the excess over \$500,000

Head of Household (2018)	
If taxable income is:	The tax is:
Not over \$13,600	10% of the taxable income
Over \$13,600 but not over \$51,800	\$1,360 plus 12% of the excess over \$13,600
Over \$51,800 but not over \$82,500	\$5,944 plus 22% of the excess over \$51,800
Over \$82,500 but not over \$157,500	\$12,698 plus 24% of the excess over \$82,500
Over \$157,500 but not over \$200,000	\$30,698 plus 32% of the excess over \$157,500
Over \$200,000 but not over \$500,000	\$44,298 plus 35% of the excess over \$200,000
Over \$500,000	\$149,298 plus 37% of the excess over \$500,000

Married Filing Jointly (2018)	
If taxable income is:	The tax is:
Not over \$19,050	10% of the taxable income
Over \$19,050 but not over \$77,400	\$1,905 plus 12% of the excess over \$19,050
Over \$77,400 but not over \$165,000	\$8,907 plus 22% of the excess over \$77,400
Over \$165,000 but not over \$315,000	\$28,179 plus 24% of the excess over \$165,000
Over \$315,000 but not over \$400,000	\$64,179 plus 32% of the excess over \$315,000
Over \$400,000 but not over \$600,000	\$91,379 plus 35% of the excess over \$400,000
Over \$600,000	\$161,379 plus 37% of the excess over \$600,000

Married Filing Separately (2018)	
If taxable income is:	The tax is:
Not over \$9,525	10% of the taxable income
Over \$9,525 but not over \$38,700	\$952.50 plus 12% of the excess over \$9,525
Over \$38,700 but not over \$82,500	\$4,453.50 plus 22% of the excess over \$38,700
Over \$82,500 but not over \$157,500	\$14,089.50 plus 24% of the excess over \$82,500
Over \$157,500 but not over \$200,000	\$32,089.50 plus 32% of the excess over \$157,500
Over \$200,000 but not over \$300,000	\$45,689.50 plus 35% of the excess over \$200,000
Over \$300,000	\$80,689.50 plus 37% of the excess over \$300,000

Example 2.5.6 Suppose Adira is filing as head of household and has a taxable income of \$86,450. Calculate her taxes using the individual tax brackets and with the simplified table.

Solution. We will calculate Adira's taxes first the long way:

$$0.10(\$13,600) + 0.12(\$51,800 - \$13,600) + 0.22(\$82,500 - \$51,800) + 0.24(\$86,450 - \$82,500)$$

$$\begin{aligned} &= 0.10(\$13,600) + 0.12(\$38,200) + 0.22(\$30,700) + 0.24(\$3,950) \\ &= \$1,360 + \$4,584 + \$6,754 + \$948 \\ &= \$12,698 + \$948 \\ &= \$13,646 \end{aligned}$$

Now, using the simplified tax table for single filing status, we see that Adira's taxable income puts her in the 24% tax bracket. The simplified table tells us that her taxes will be equal to \$12,698 plus 24% of the excess over \$82,500. Notice the number \$12,698 is the same value we got for all of the "full" tax brackets, so we only need to calculate the last one. To find the excess over \$82,500, we subtract \$82,500 from her taxable income. Thus, her taxes are:

$$\begin{aligned} \$12,698 + 0.24(\$86,450 - \$82,500) &= \$12,698 + 0.24(\$3,950) \\ &= \$12,698 + \$948 \\ &= \$13,646 \end{aligned}$$

We see that both methods result in the same value, \$13,646, for Adira's taxes. □

Example 2.5.7 Phyllis and Gladys are married and filing jointly. Together their taxable income is \$112,000. Use the simplified 2018 tax tables from this section to determine how much they owe in taxes.

Solution. Since Phyllis and Gladys are married and filing jointly, their taxable income puts them in the 22% tax bracket. Using the simplified 2018 tax table, their taxes are \$8,907 plus 22% of the excess over \$77,400:

$$\begin{aligned} \$8,907 + 0.22(\$112,000 - \$77,400) &= \$8,907 + 0.22(\$34,600) \\ &= \$8,907 + \$7,612 \\ &= \$16,519 \end{aligned}$$

Phyllis and Gladys owe \$16,519 in taxes. □

2.5.12 Tax Credits

Tax credits are different from deductions in that they reduce the amount of tax you owe by the full amount of the credit, not just a percentage. This makes credits much more valuable than deductions.

Common tax credits include child tax credits, earned income credits, child and dependent care credits, American opportunity tax credits, lifetime learning credits, and various federal energy credits.

Example 2.5.8 Shiro is in the 22% tax bracket and itemizes his deductions. How much will his tax bill be reduced if he makes a $1,000 contribution to charity? How much will his bill be reduced if he gets a $1,000 tax credit?

Solution. The tax credit will reduce his tax bill by the full amount of the credit, so the $1,000 tax credit will reduce his tax bill by $1,000. As a deduction, however, his contribution to charity will only reduce his tax bill by 22% of the $1,000, or 0.22($1,000) = $220. Tax credits are always better than deductions. □

2.5.13 Calculating a Refund or Payment Due

Employers are required to take out an estimated amount for taxes from each of our paychecks. These withholdings are taken out, so we don't all have huge tax bills at the end of the year, and so the government has the income it needs to run. When you file your taxes, you compare the amount of tax you actually owe, with the withholdings from your paycheck. If you had more withheld during the year than you owe, you will file for a refund. If your withholdings were less, though, you must pay the difference. Let's look at a couple of examples with tax credits and withholdings.

Example 2.5.9 John's taxes are $4,342.50. He can claim an American opportunity tax credit of $2,300 and he had $4,135 withheld from his paychecks. Determine if John will owe money or get a refund.

Solution. The first step is to reduce John's taxes by the full amount of the tax credit.

$\$4,342.50 - \$2,300 = \$3,042.50$

Next we subtract the amount withheld from his paychecks by his employer. Since his withholdings are greater than the amount he owes after applying the tax credit, he will receive a refund equal to the difference.

$\$3,042.50 - \$4,135 = -\$1,092.50$

John will receive a refund of $1,092.50. Notice that having a negative amount after subtracting credits and withholdings means you will receive a refund, while having a positive amount means you still owe money. □

Example 2.5.10 As we discovered in Example 2.5.7, Phyllis and Gladys owe $16,519 in taxes. Their employers withheld $8,980 and they received a $7,500 credit for their purchase of an electric car. Will they receive a refund, or will they need to make a payment?

Solution. To determine whether they will receive a refund or if they will still owe money, we will first subtract the full amount of the electric car credit.

$\$16,519 - \$7,500 = \$9,019$

We now subtract their withholdings:

$\$9,019 - \$8,980 = \$39$

Since the value is positive, Phyllis and Gladys still owe $39 in taxes. □

2.5.14 Exercises

1. Which decreases your tax bill more, a credit or a deduction?

2. You are in the 12% tax bracket and get a credit of $500. How has the amount of taxes owed changed?

3. You are in the 12% tax bracket and get a deduction of $500. How has the amount of taxes owed changed?

4. Can you take the standard deduction and itemize your deductions?

5. Can you make adjustments to your income and take the standard deduction?

6. If you decide to take the standard deduction what have you considered?

7. Shaysiah has paid $11,500 in mortgage interest which she can take as a deduction and has $1,500 in other deductions that she can take this year. She is filing as single. Should she itemize her deductions or take the standard deduction?

8. If you are married do you have to file your taxes together?

9. Fredrick is concerned about the effect of a raise on his taxes. He's getting a raise of $3,000, putting him into a higher tax bracket by $1,000 dollars. He's concerned about his entire income being taxed at 22% instead of 12%. Should he be concerned? Why or why not?

10. A single person with taxable income of $95,000 will have the first $9,525 of that income taxed at what rate? Determine the taxes owed on just the first $9,525.

11. A single person with taxable income of $185,000 will have the first $9,525 of that income taxed at a different rate than the income between $9,525 and $38,700. Determine the taxes owed on just the second bracket.

12. Using the simplified 2018 tax table, determine the income tax owed for a single person who has $80,000 of taxable income.

13. Doug and Chris are filing jointly. They owe $16,589 in taxes. Throughout the year they had $13,456 withheld from their paychecks, and they can claim an energy credit of $2,500 for purchasing a hybrid vehicle. Determine the amount they will need to pay or will get refunded.

14. Heather owes $7,589 in taxes. Throughout the year she had $6,456 withheld from her paychecks and she can claim an education credit of $1,980. Determine the amount she will need to pay or will get refunded.

15. Determine the amount of taxes owed or the refund that would result in this situation:

- Filing Status: Married filing jointly
- Gross Income: $125,000
- Adjustments: $5,600
- Itemized Deductions: $11,400
- Credits: $15,000

  a. What is their adjusted gross income (AGI)?
  b. Should they itemize or take the standard deduction?
  c. Use the simplified 2018 tax tables to determine their taxes.
  d. What is their final tax refund or amount still owed?

16. Francis and Edward are planning to get married and they want to determine whether there is an advantage or disadvantage to marrying before the end of the year and filing their taxes jointly. Use the information below to calculate the amount they would owe or receive if they each filed as single, and the amount they would owe or receive if they filed jointly as a married couple. Which is the better choice?

- Filing Status: TBD
- Francis' Gross Income: $35,000
- Edward's Gross Income: $40,000

- Francis' Adjustments: \$7000
- Edward's Adjustments: \$3000
- Francis' Withholdings: \$14000
- Edward's Withholdings: \$5500
- Francis' Credits: \$4000
- Edward's Credits: \$5000

17. Janice is unmarried and has two kids. She earned \$76,000 in wages last year, received \$750 in interest from a savings account, and contributed \$25,000 to a tax deferred savings account. Her itemized deductions are \$19,600.

a. Determine Janice's gross income.

b. Determine Janice's adjusted gross income.

c. Should Janice take the standard deduction or itemize? Explain.

d. Determine Janice's taxable income.

e. If Janice has \$4,200 in child tax credits and had \$6,300 withheld for taxes from her wages, will Janice owe money, or will she receive a refund? Calculate the amount.

2.6 Chapter 2 Review

Review Exercises

1. If you deposit \$1,525 in an account at 5.6% APR for fourteen years, how much will you have in the account and how much interest did you earn after 14 years if:

a. The interest is compounded using simple interest.

b. The interest is compounded quarterly.

c. The interest is compounded weekly.

d. The interest is compounded continuously.

2. Alicia takes out a \$2,300 loan that charges 15% APR simple interest. How much will she pay if she has the loan for 3 years? How much interest did she pay?

3. Devon invests \$25,000 into an account for 15 years. He earns 6.5% interest compounded quarterly. What is the future value and how much interest did he earn?

4. The current inflation rate is 2.92%. If this continues for the next 10 years find the cost of the following items in the year 2029. Note: Inflation is Continuous Compound Interest.

a. Gas Average: \$3.49

b. Dozen Eggs: \$1.99

c. Bread: \$3.29

d. Basic Monthly Cell Phone bill: \$79.99

e. Average cost of downloading a song: \$1.99

5. You are purchasing new furniture that costs \$3500. You are required to make a down payment of \$350. The loan will be a simple interest at 13% APR and the length of the loan will be 28 months. What is your monthly payment and how much did you pay back?

6. Tom has misplaced the sales contract for his car and cannot remember the amount he originally financed. He does know that the interest rate was 9.6% APR for 60 months and the simple interest loan required a total of 60 payments at \$254.23. What is the amount of money that Tom borrowed?

7. For each find the Future Value, the total amount deposited, and the interest earned.

a. Regular Deposit \$350, Compounded Monthly, 6.5% APR for 25 years

b. Regular Deposit \$500, Compounded Quarterly, 6.5% APR for 15 years

c. Regular Deposit \$75, Compounded Weekly, 4.5% APR for 30 years

8. For each house find the monthly payment for a 30 year loan and 5% APR. Find the amount of interest you pay on each loan.

- House 1: John's Landing Townhouse: \$299,900
- House 2: Gresham 4 bedroom House: \$389,900

9. Find the net monthly cash flow (1 month = 4 weeks)

Income:	Expense
Job Income: \$475 per week	Rent: \$650 per month
Loan: \$2500 per term. (10 weeks)	Groceries: \$55 per week
	Tuition and fees: \$3000 per term
	Books: \$255 per term
	Miscellaneous: \$75 per week

10. Pat loves painting. With tax, the total spent is about \$46 each month on supplies. Once a week (52 weeks per year) the art class cost Pat \$15.75. How much is Pat spending on painting in a year?

11. If you are in the 12% tax bracket and can take a \$1,000 deduction, how much will your tax bill decrease by?

12. If you are in the 12% tax bracket and can take a \$1,000 credit, how much will your tax bill decrease by?

13. Amir's made \$43,000 in wages and \$1000 in tips. He contributed \$3,000 into his IRA account.

a. Find Amir's Gross Income.

b. Find Amir's Adjusted Gross Income (AGI).

14. Amir has \$9,540 he could take in itemized deductions. The standard deduction for a single filer is \$12,000.

a. Find Amir's taxable income.

b. Use the 2018 tax table determine to determine how much Amir owes in taxes.

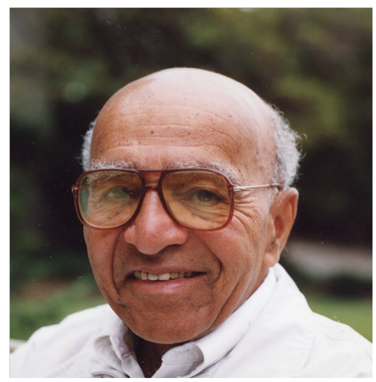
15. Amir can take a \$1200 education credit and has had \$2700 withheld from his paychecks.

a. Determine the amount Amir will owe or be refunded.

Chapter 3

Statistics

Dr. David Blackwell and Statistics.



Dr. David Blackwell[1] (1919 – 2010) was a theoretical statistician noted for his teaching and work in game theory and probability theory. He was the first and only African American member of the National Academy of Sciences (Williams, 2008). Blackwell's research in mathematics and statistics have found application in many fields including economics and accounting. Dr. Blackwell grew up in Centralia, Ill. In school, he was intrigued by geometry and calculus. His high school math club advisor would challenge members with problems from the School Science and Mathematics journal and submit their solutions. Blackwell was identified three times in the magazine as having solved problems and one of his solutions was published (David Harold Blackwell: National Visionary Leadership Project: African American History, 2013).

At 16, Blackwell enrolled at the University of Illinois and majored in mathematics. After three years, he graduated and continued at the university to obtain his master's and doctorate degrees. Dr. Blackwell was given a one-year appointment as a Rosenwald Postdoctoral Fellow at the Institute for Advanced Study at Princeton University in 1941. It was common for Institute members to be made visiting fellows of Princeton. This created a controversy because there were no Black students enrolled at the university during that time. Princeton's president wrote a letter to the Institute protesting Blackwell's admission, but the Institute honored the appointment (David Harold Blackwell: National Visionary Leadership Project: African American History, 2013).

When his fellowship was drawing to a close, Dr. Blackwell applied for teaching positions at 105 Black colleges. He didn't apply to white institutions because he assumed they would not accept him because of his race (Williams, 2008). His first teaching job was at Southern University in Baton Rouge, La.

Dr. Blackwell's focus later shifted to the field of statistics and is famous for the Rao-Blackwell Theorem. In 1954, he accepted a professorship at the University of California at Berkeley, where he was previously unable to teach due to racism. By 1956, he was appointed chairman of the statistics department. He continued teaching and publishing a substantial amount of research until his retirement as professor emeritus in 1989. He is the recipient of numerous awards and holds honorary degrees from 12 universities, among them Carnegie-Mellon, Yale, Howard and Harvard.

[1] Photo: "David Blackwell" by George M. Bergman is licensed under CC BY-SA 4.0.

3.1 Overview of the Statistical Process

Objectives: Section 3.1 Overview of the Statistical Process

Students will be able to:

- Define and identify the population, parameter, sample and statistic
- Identify four sampling methods: simple random sample (SRS), stratified, systematic and convenience
- Identify and discuss types of bias association with sampling
- Distinguish between experimental and observational studies
- Explain margin of error and confidence intervals

Introduction to Statistics. We are bombarded by information and statistics every day. But if we cannot distinguish credible information from misleading information, then we are vulnerable to manipulation and making decisions that are not in our best interest. Statistics provides tools for us to evaluate information critically. In this sense, statistics is one of the most important things to know about.

Statistics are often presented to add credibility to an argument. To give some examples, here are some claims that we have heard on several occasions. (We are not saying that each one of these claims is true!)

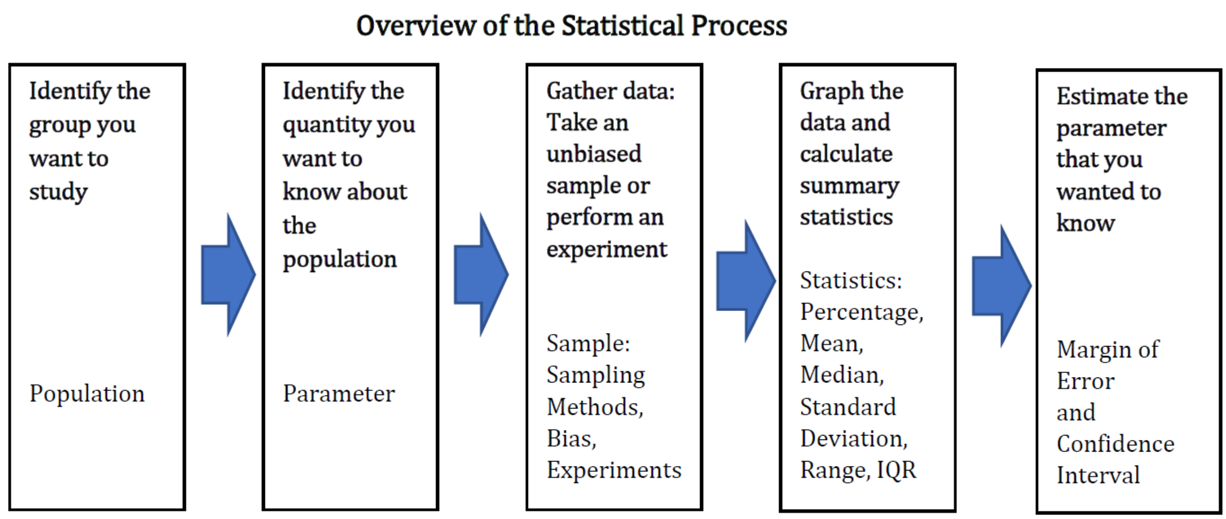
- 4 out of 5 dentists recommend Dentyne.
- Almost 85% of lung cancers in men and 45% in women are tobacco-related.
- Condoms are effective 94% of the time.
- People tend to be more persuasive when they look others directly in the eye and speak loudly and quickly.
- Women make 75 cents to every dollar 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 these claims are statistical in character. We suspect that some of them sound familiar; if not, you have probably heard other claims like them. Notice how diverse the examples are; they come from psychology, health, law, sports, business, etc. Data and data-interpretation show up in virtually every facet of contemporary life.

Many of these numbers do not represent careful statistical analysis. They can be misleading and push you into decisions that you might regret. This chapter will help you learn the skills to be a critical consumer of statistical claims.

3.1.1 Statistical Process

To give you an overview, here is a diagram of the steps taken in a poll or other statistical study, and the elements in each step that we will discuss in this chapter. We will use many examples to illustrate the whole process.



3.1.2 Population

Before we begin gathering any data to analyze, we need to identify the population we are studying. The population of a study is the group we want to know something about. The ***population*** could be people, auto parts or tomato plants.

If we want to know the amount of money spent on textbooks by a typical college student, our population might be all students at Portland Community College. Or it might be:

1. All community college students in the state of Oregon.
2. All students at public colleges and universities in the state of Oregon.
3. All students at all colleges and universities in the state of Oregon.
4. All students at all colleges and universities in the entire United States.
5. And so on.

The intended population is also called the ***target population***, since if we design our study badly, the collected data might not actually be representative of the intended population.

Example 3.1.1 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. □

3.1.3 Parameter

A ***parameter*** is the value (percentage, average, etc.) that we are interested in for the whole population. Since it is often too time-consuming, expensive and/or impossible to get data for the entire population, the parameter is usually a theoretical quantity that we are trying to estimate. For example, the typical amount of money spent per year on textbooks by students at your college in a year is a parameter.

3.1.4 Sample

To estimate the parameter, we select a ***sample***, which is a smaller subset of the entire population. It is very important that we choose a ***representative sample***, one that matches the characteristics of the population, to have a good estimate. If we survey 100 students at your college, those students would be our sample. We will talk about how to choose a representative sample later in this section.

3.1.5 Statistic

To get our ***data***, we would then ask each student in the sample how much they spent on textbooks and record the answers, or ***raw data***. Then we could calculate the average, which is our statistic. A ***statistic*** is a value (percentage, average, etc.) calculated using data from a sample.

Example 3.1.2 A researcher wants to know how the citizens of Portland feel about a voter initiative. To study this, they go downtown to the Pioneer Place Mall and survey 500 shoppers. Sixty percent indicate they are supportive of the initiative. Identify the population, parameter, sample and statistic.

Solution. *Population*: While the intended population of this survey is Portland citizens, the effective population is Pioneer Place Mall shoppers. There is no reason to assume that shoppers at this mall would be representative of all Portland citizens.

Parameter: The parameter is what we want to know about the population, the percentage of Portland citizens that support the initiative.

Sample: The sample is the subgroup of the population selected. The 500 shoppers questioned make up the sample, which, again, is probably not representative of the population.

Statistic: The statistic is calculated using the data from the sample. The percentage of people sampled who support the initiative is 60%. □

3.1.6 Sampling

As we mentioned in a previous section, the first thing we should do before conducting a survey is to identify the population that we want to study. Suppose we are hired by a politician to determine the amount of support they have among the electorate should they decide to run for another term. What population should we study? Every person in the district? Eligible voters might be better, but what if they don't register? Registered voters may not vote. What about "likely voters?"

This is the criteria used in a lot of political polling, but it is sometimes difficult to define a "likely voter." Here is an example of the challenges of political polling.

Example 3.1.3 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 that 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 lower turnout rate for elections) 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. □

So, identifying the population can be a difficult job, but once we have identified the population, how do we choose a good sample? We want our statistic to estimate the parameter we are interested in, so we need to have a representative sample. Returning to our hypothetical job as a political pollster, we would not anticipate very accurate results if we drew all of our samples from customers at a Starbucks, or your list of TikTok followers. How do we get a sample that resembles our population?

3.1.7 Sampling Methods

One way to get a representative sample is to use randomness. We will look at three types of sampling that use *randomness* and one that does not.

3.1.8 Simple random sample (SRS)

A ***simple random sample***, abbreviated SRS, is one in which every member of the population has an equal probability of being chosen.

Example 3.1.4 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. 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 Oregon, some of these samples might tend to have a slightly higher percentage of Democrats (or Republicans) than the general population; some samples might include more older people and some samples might include more younger people; etc. This is called ***sampling variation***. If there are certain groups that we want to make sure are represented, we might instead use a stratified sample.

3.1.9 Stratified sampling

In ***stratified sampling***, a population is divided into a number of subgroups (or strata). Random samples are then taken from each subgroup. It is often desirable to make the sample sizes proportional to the size of each subgroup in the population.

Example 3.1.5 Suppose that data from voter registrations in the state 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. □

A way to remember stratified sampling is think about having a piece of layer cake. Each layer represents a stratum or subgroup, and a slice of the cake represents a sample of each layer.

3.1.10 Systematic sampling

In ***systematic sampling***, every n^{th} member of the population is selected to be in the sample. The starting position is often chosen at random.

Example 3.1.6 To select a systematic sample, Portland Community College could use their database to select a random student from the first 100 student ID numbers. Then they would select every 100th student ID number after that. Systematic sampling is not as random as a simple random sample (if your ID number is right next to your friend's because you applied at the same time, you could not both end up in the same sample) but it can yield acceptable samples. This method can be useful for people waiting in lines, parts on a manufacturing line, or plants in a row. □

3.1.11 Convenience sampling

Convenience sampling is when samples are chosen by selecting whomever is convenient. This is the worst type of sampling because it does not use randomness.

Example 3.1.7 A pollster stands on a street corner and interviews the first 100 people who agree to speak to them. This is a convenience sample. □

3.1.12 Statistical Bias

There is no way to correct for biased data, so it is very important to think through the entire study and data analysis before we start. We talked about ***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 can lead to bias if the people who volunteer have different characteristics than the general population. Here is a summary of some additional sources of bias.

3.1.13 Types 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 made up of 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 refuse to participate in a study or drop out of an experiment, we can no longer be certain that our sample is representative of the population

Sources of bias may be conscious or unconscious. They may be innocent or as intentional as pressuring by a pollster. Here are some examples of the types of bias.

Example 3.1.8

a. 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. This is an example of a self-interest study; one in which the researches have a vested interest in the outcome of the study. While this does not necessarily mean the study was biased, we should subject the study to extra scrutiny.

b. Consider online reviews of products and businesses. Customers tend to leave reviews if they are very satisfied or very dissatisfied. While you can look for overall patterns and get useful information, these reviews suffer from voluntary response bias and likely capture more extreme views than the general population.

c. A survey asks participants a question about their interactions with people of different ethnicities. This study could suffer from response bias. A respondent might give an untruthful answer to not be perceived as racist.

d. An employer puts out a survey asking their employees if they have a drug abuse problem and need treatment help. Here, answering truthfully might have serious consequences; responses might not be accurate if there is a perceived lack of anonymity and employees fear retribution.

e. A survey asks, "Do you support funding research on alternative energy sources to reduce our reliance on high-polluting fossil fuels?" This is an example of a loaded or leading question – questions whose wording leads the respondent towards a certain answer.

f. A poll was conducted by phone with the question, "Do you often have time to relax and read a book?" Fifty percent of the people who were called refused to participate in the survey (Probably because they didn't have the time). It is unlikely that the results will be representative of the entire population. This is an example of non-response bias.

□

Loaded questions can occur intentionally by pollsters with an agenda, or accidentally through poor question wording. Also of concern is question order, where the order of questions changes the results. Here is an example from a psychology researcher[2]:

Example 3.1.9 *"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."* □

3.1.14 Observational Studies

So far, we have primarily discussed surveys and polls, which are types of ***observational studies*** – studies based on observations or measurements. These observations may be solicited, like in a survey or poll. Or, they may 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.

3.1.15 Experiments

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. When conducting experiments, it is essential to isolate the treatment being tested. Here are some examples of treatments.

Example 3.1.10

a. 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.

b. 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.

c. A psychology researcher explores the effect of music on affect by measuring people's mood while listening to different types of music. The music is the treatment.

---

[1] Reuters. http://news.yahoo.com/s/nm/20090423/od_uk_nm/oukoe_uk_gum_learning. Retrieved 4/27/09

[2] Swartz, Norbert. http://www.umich.edu/~newsinfo/MT/01/Fal01/mt6f01.html Retrieved 3/31/2009

d. Suppose a middle school finds that their students are not scoring well on the state's standardized math test. They decide to run an experiment to see if a new 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 the last scenario is that it is not clear whether the new curriculum or the math specialist is responsible for the improvement. This is called confounding and it is the downfall of many experiments, though it is often hidden.

□

3.1.16 Confounding

Confounding occurs when there are two or more potential variables that could have caused the outcome and it is not possible to determine which one actually caused the result.

Example 3.1.11

a. 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.

b. 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.

□

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.

3.1.17 Control group

In experiments, the participants are typically divided into a treatment group and a control 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 each gender identity, 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.

Example 3.1.12

a. 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 took the SAT test.

b. A company testing a new plant food grows two crops of plants in adjacent fields that typically produce the same amount of food. The treatment group receives the new plant food and the control group does not. The crop yields would then be compared. By growing the two crops at the same time in similar 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, or a response to have something. This is called a placebo effect.

3.1.18 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 3.1.13 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.

3.1.19 Placebo and Placebo-controlled experiments

An experiment that gives the control group a placebo is called a ***placebo-controlled*** experiment.

Example 3.1.14

a. In a study for a new medicine that is dispensed in a pill form, a sugar pill could be used as a placebo.

b. In a study on the effect of alcohol on memory, a non-alcoholic beer might be given to the control group as a placebo.

c. In a study of a frozen meal diet plan, the treatment group would receive the diet food, and the control group could be given standard frozen meals taken out of their original packaging.

□

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 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. It is very important to consider the ethical ramifications of any experiment.

3.1.20 Blind studies

A ***blind study*** is one that uses a placebo and the participants do not know whether they are receiving the treatment or a placebo. A ***double-blind study*** is one in which the subjects and those interacting with them don't know who is in the treatment group and who is in the control group.

Example 3.1.15 In a study about anti-depression medication, you would not want the psychological evaluator to know whether a patient is in the treatment or control group, as it might influence their evaluation. The experiment should be conducted as a double-blind study. □

[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.

3.1.21 Margin of Error and Confidence Intervals

Even when a study or experiment has successfully avoided bias and has been well done, there is still an element of variation. If we took 5 different random samples of 100 college students and calculated their average textbook cost, we wouldn't expect to get the exact same average for each sample. This is due to sampling variation. To account for this, researchers publish their margin of error or a confidence interval for their statistics. These numbers describe the precision of the estimate for a certain confidence level.

You've probably heard something like, "The candidate has 54 percent of the likely voters, plus or minus three percent." The 3% is called the ***margin of error***, so the true percentage is somewhere between 51% and 57%, with a certain level of confidence. To write this as a ***confidence interval***, we place the numbers in parentheses from smallest to largest, separated by a comma: (51%, 57%).

The most common ***confidence level*** is 95%, which means if the poll was conducted repeatedly, and we made a confidence interval each time, we would expect the true percentage, or parameter, to fall within our confidence interval 95 out of 100 times. You can learn more on how to calculate the margin of error for different confidence levels in a statistics class.

Example 3.1.16 Let's say we asked a random sample of 100 students at Portland Community College and found that they spent an average of $451.32 on books their first year, plus or minus $85.63. Write this as a confidence interval, assuming a 95% confidence level.

Solution. If the margin of error was calculated to be plus or minus $85.63, then with a confidence level of 95% we could say that the average amount spent by the population is somewhere between $365.69 and $536.95.

These values are found by computing $\$451.32 - \$85.63 = \$365.69$ and $\$451.32 + \$85.63 = \$536.95$

We could also write this as a ***confidence interval***: ($365.69, $536.95).

Now we have come full circle and seen how we can use data from a sample to estimate the parameter we were interested in for our population. □

3.1.22 Exercises

1. Describe the difference between a sample and a population.

2. Describe the difference between a statistic and a parameter.

3. The ASPCC randomly selects 200 students from PCC Cascade campus to participate in a childcare survey in order to determine the demand for additional childcare options for PCC students.

a. Who is the intended population?

b. What is the sample?

c. Is the collected data representative of the intended population? Why or why not?

4. A local research firm randomly selects 1200 homes in Washington County to determine support for adding compost pick up to residents' regular garbage service.

a. Who is the intended population?

b. What is the sample?

c. Is the collected data representative of the intended population? Why or why not?

5. 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. Who 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 statistic for the percentage of representatives surveyed who said they were supporting the education bill.

e. If the margin of error was 5%, give the confidence interval for the percentage of representatives we might we expect to support the education bill and explain what the confidence interval tells us.

6. The city of Raleigh has 9,500 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. Who 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 statistic for the percentage of voters surveyed who said they'd vote for Brown.

e. If the margin of error was 3.5%, give the confidence interval for the percentage of voters surveyed that we might we expect to vote for Brown and explain what the confidence interval tells us.

7. To determine the average length of trout in a lake, researchers catch 20 fish and measure them. Describe the population and sample of this study.

8. To determine the average diameter of evergreen trees in a forested park, researchers randomly tag 45 specimens and measure their diameter. Describe the population and sample of this study.

9. A college reports that the average age of their students is 28 years old. Is this a parameter or a statistic?

10. A local newspaper reports that among a sample of 250 subscribers, 45% are over the age of 50. Is this a parameter or a statistic?

11. A recent survey reported that 64% of respondents were in favor of expanding the BIKETOWN bike share system to the greater Portland area. Is this a parameter or a statistic?

12. Which sampling method is being described?

a. In a study, the sample is chosen by separating all cars by size and selecting 10 of each size grouping.

b. 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.

c. Every 4th person on the class roster was selected.

13. Which sampling method is being described?

a. A sample was selected to contain 25 people aged 18-34 and 30 people aged 35-70.

b. Viewers of a new show are asked to respond to a poll on the show's website.

c. To survey voters in a town, a polling company randomly selects 100 addresses from a database and interviews those residents.

14. Identify the most relevant source of bias in each situation.

a. A survey asks the following: Should the mall prohibit loud and annoying rock music in clothing stores catering to teenagers?

b. To determine opinions on voter support for a downtown renovation project, a surveyor randomly questions people working in downtown businesses.

c. A survey asks people to report their actual income and the income they reported on their IRS tax form.

d. A survey randomly calls people from the phone book and asks them to answer a long series of questions.

e. The Beef Council releases a study stating that consuming red meat poses little cardiovascular risk.

f. A poll asks, "Do you support a new transportation tax, or would you prefer to see our public transportation system fall apart?"

15. Identify the most relevant source of bias in each situation.

a. A survey asks the following: Should the death penalty be permitted if innocent people might die?

b. A study seeks to investigate whether a new pain medication is safe to market to the public. They test by randomly selecting 300 people who identify as men from a set of volunteers.

c. A survey asks how many sexual partners a person has had in the last year.

d. A radio station asks listeners to phone in their response to a daily poll.

e. A substitute teacher wants to know how students in the class did on their last test. The teacher asks the 10 students sitting in the front row to state their latest test score.

f. High school students are asked if they have consumed alcohol in the last two weeks.

16. Identify whether each situation describes an observational study or an experiment.

a. The temperature on randomly selected days throughout the year was measured.

b. One group of students listened to music and another group did not while they took a test and their scores were recorded.

c. The weights of 30 randomly selected people are measured.

17. Identify whether each situation describes an observational study or an experiment.

a. Subjects are asked to do 20 jumping jacks, and then their heart rates are measured.

b. Twenty coffee drinkers and twenty tea drinkers are given a concentration test.

c. The weights of potato chip bags are weighed on the production line before they are put into boxes.

18. A team of researchers is testing the effectiveness of a new vaccine for human papilloma virus (HPV). 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?

19. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment. Suppose that a new cancer treatment is under study. 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 cancer patients throughout this new treatment.

a. What is the population of this study?

b. Would you expect the data from the two researchers to be identical? Why or why not?

c. If the first researcher collected their data by randomly selecting 10 nearby ZIP codes, then selecting 4 people from each, which sampling method did they use?

d. If the second researcher collected their data by choosing 40 patients they knew, what sampling method did they use? What concerns would you have about this data set, based upon the data collection method?

20. For the clinical trials of a weight loss drug containing *Garcinia Cambogia* the 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?

21. A study is conducted to determine whether people learn better with routine or crammed studying. Subjects volunteer from an introductory psychology class. At the beginning of the semester 12 subjects volunteer and are assigned to the routine studying group. At the end of the semester 12 subjects volunteer and are assigned to the crammed studying group.

a. Identify the target population and the sample.

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 a specified population and 2. Subjects are not randomly assigned to groups. Which problem is more serious? What effect on the results does each have?

22. 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. The second group is asked to tell the truth on the first half of the questions and lie on the second half. 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? Explain.

23. A poll found that 30%, plus or minus 5% of college freshmen prefer morning classes to afternoon classes.

a. What is the margin of error?

b. Write the survey results as a confidence interval.

c. Explain what the confidence interval tells us about the percentage of college freshmen who prefer morning classes?

24. A poll found that 38% of U.S. employees are engaged at work, plus or minus 3.5%.

a. What is the margin of error?

b. Write the survey results as a confidence interval.

c. Explain what the confidence interval tells us about the percentage of U.S. employees who are engaged at work?

25. A recent study reported a confidence interval of (24%, 36%) for the percentage of U.S. adults who plan to purchase an electric car in the next 5 years.

a. What is the statistic from this study?

b. What is the margin of error?

26. A recent study reported a confidence interval of (44%, 52%) for the percentage of two-year college students who are food insecure.

a. What is the statistic from this study?

b. What is the margin of error?

27. 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.

28. A sports psychologist believes that people are more likely to be extroverted as an adult if they played team sports as a child. Describe two possible studies to test this theory. Design one as an observational study and the other as an experiment. Which is more practical?

29. Find a newspaper or magazine article, or the online equivalent, describing the results of a recent study (not a simple poll). 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, margin of error, etc.

30. Use a polling website such as www.pewresearch.com or www.gallup.com and search for a poll that interests you. Find the result, the margin of error and confidence level for the poll and write the confidence interval.

3.2 Describing Data

Objectives: Section 3.2 Describing Data

Students will be able to:

- Define and identify categorical and quantitative data
- Read and construct frequency tables and relative frequency tables
- Make bar charts and pie charts for categorical variables by hand and/or using technology
- Identify elements of misleading graphs: 3-dimensional graphs, perceptual distortion, misleading scales, stacked bar graphs
- Make histograms for quantitative variables by hand and/or using technology
- Identify the number of modes in a distribution and whether it is zymmetric, skewed to the left, or skewed to the right

Once we have collected data from an observational study or an experiment, we need to summarize and present it in a way that will be meaningful to our audience. The raw data is not very useful by itself. In this section we will begin with graphical presentations of data and in the rest of the chapter we will learn about numerical summaries of data.

3.2.1 Types of Data

There are two types of data, categorical data and quantitative data.

Categorical (qualitative) data are pieces of information that allow us to classify the subjects into various categories.

Example 3.2.1 We might conduct a survey to determine the name of the favorite movie that people saw in a movie theater. When we conduct such a survey, the responses would look like: *Finding Nemo, Black Panther, Titanic, etc.*

We can 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 *"Black Panther"* because it is categorical data. □

Quantitative data are responses that are numerical in nature and with which we can perform meaningful calculations.

Example 3.2.2 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, ...). 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, 110 minutes, etc.) or the amount of money you paid for a movie ticket the last time you went to a movie theater ($5.50, $9.75, $10.50, etc.).

We cannot assume that all numbers are quantitative data, and sometimes it is not so clear-cut. Here are some examples to illustrate this.

Example 3.2.3

a. Suppose we gather respondents' ZIP codes in a survey to track their geographical location. ZIP codes are numbers, but we can't do any meaningful 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.

b. A survey about the movie you most recently saw includes the question, "How would you rate the movie?" with these possible answers:

1 It was awful.

2 It was just okay.

3 I liked it.

4 It was great.

5 Best movie ever!

Again, there are numbers associated with the responses, but these are really categories. A movie that rates a 4 is not necessarily twice as good as a movie that rates a 2, whatever that means; However, we often see that a movie got an average of 3.7 stars, which is an average of categorical ratings and it can give us important information. □

Overall, it is important to look at the purpose of the study for any variables that could be classified as either categorical or quantitative. Another consideration is what you plan to do with the data. Next, we will talk about how to display each type of data.

3.2.2 Presenting Categorical Data

Since we can't do calculations with categorical data, we begin by summarizing the data in a frequency table or a relative frequency table.

3.2.3 Frequency Tables

A ***frequency table*** has one column for the categories, and another for the ***frequency***, or number of times that category occurred.

Example 3.2.4 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 this table.

Car Color	Frequency of Total-Loss Collisions
Blue	25
Green	52
Red	41
White	36
Black	39
Grey	23
Total	216

□

3.2.4 Relative Frequency Tables

Numbers are usually not as easy to interpret as percentages, so we will add a column for the relative frequencies. A ***relative frequency*** is the percentage for the category, found by dividing each frequency by the total and converting to a percentage. You'll notice the percentages may not add up to exactly 100% due to rounding.

Example 3.2.5 Example 3.2.4 continued

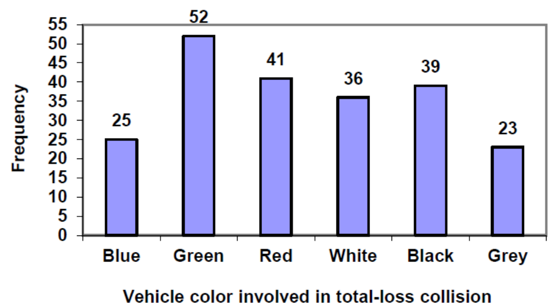
Car Color	Frequency of Total-Loss Collisions	Relative Frequency of Total-Loss Collisions
Blue	25	$25/216 = 0.116$ or 11.6%
Green	52	$52/216 = 0.241$ or 24.1%
Red	41	$41/216 = 0.190$ or 19.0%
White	36	$36/216 = 0.167$ or 16.7%
Black	39	$39/216 = 0.181$ or 18.1%
Grey	23	$23/216 = 0.107$ or 10.7%
Total	216	$216/216 = 1.0$ or 100%

It would be even more useful to have a visual to see what is going on, and this is where charts and graphs come in. For categorical data we can display our data using bar graphs and pie charts. □

3.2.5 Bar graphs

A ***bar graph*** is a graph that displays a bar for each category with the height of the bar indicating the frequency of that category. To construct a bar graph with vertical bars, we label the horizontal axis with the categories. The vertical axis will have a scale for the frequency or relative frequency.

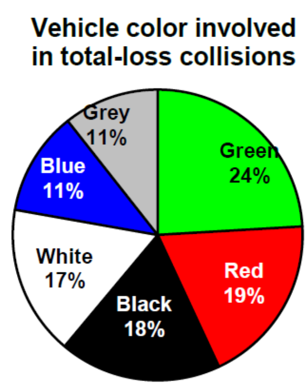
The highest frequency in our car data is 52 collisions, so we will set our vertical axis to go from 0 to 55, with a scale of 5 units. To draw bar graphs by hand graph paper is useful, or you can use technology. It is also very helpful to label each bar with the frequency or relative frequency.



3.2.6 Pie Charts

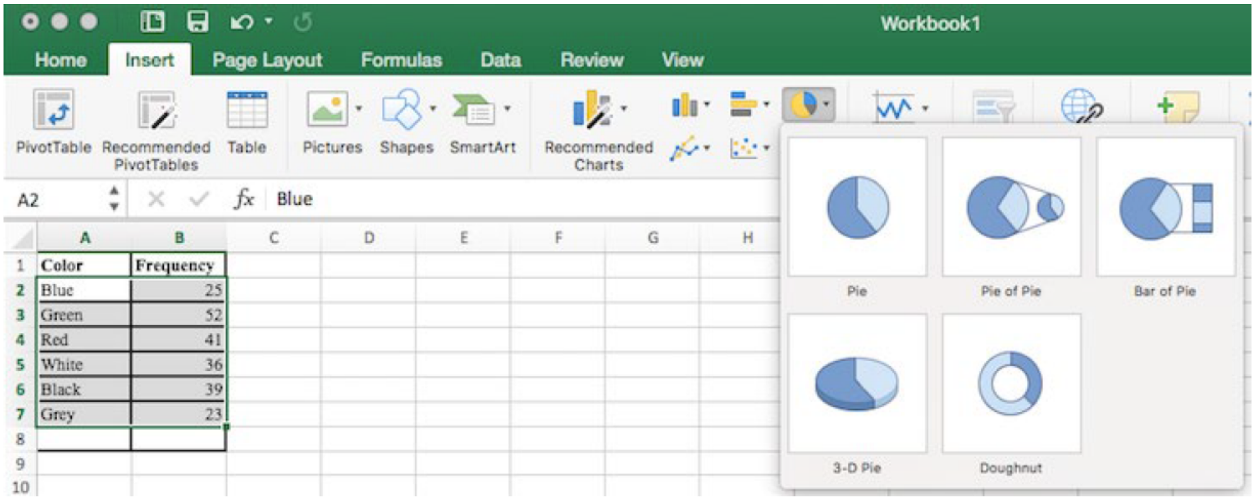
A natural way to visualize relative frequencies is with a pie chart. A ***pie chart*** is a circle with wedges cut of varying sizes like slices of pizza or pie. The size of each wedge corresponds to the relative frequency of the category. The slices add up to 100%, just like relative frequencies. Pie charts can often benefit from including frequencies or relative frequencies in the pie slices.

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. A spreadsheet is much better suited to drawing pie charts.



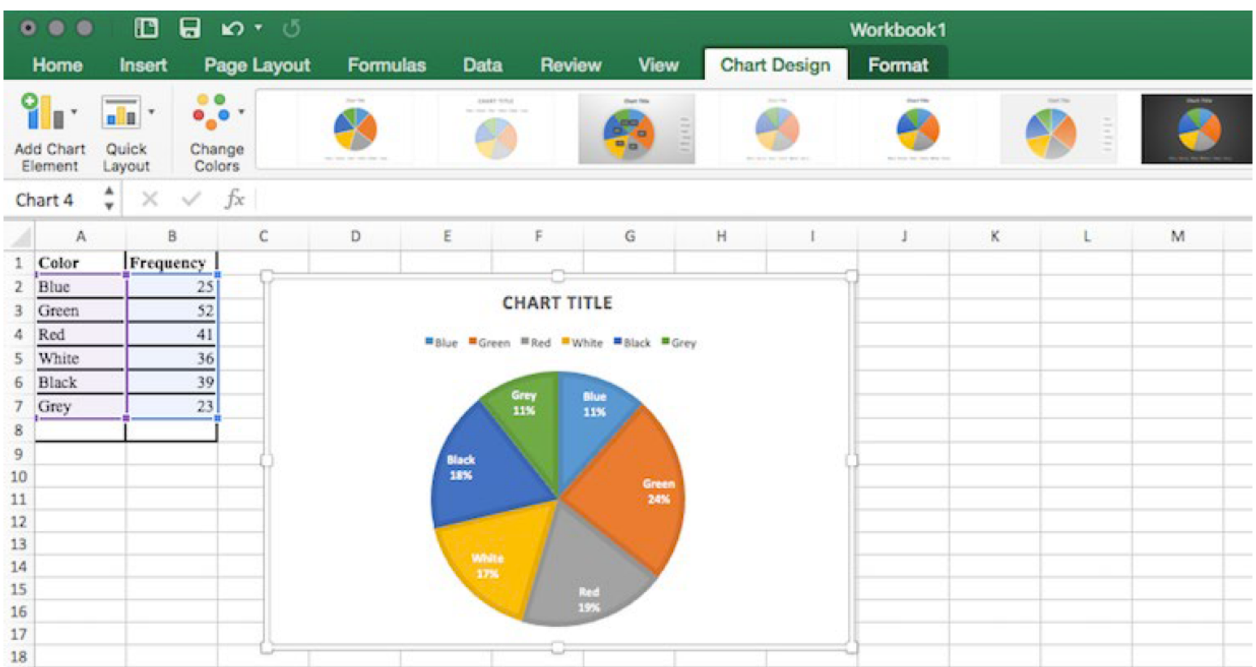
3.2.7 Using a Spreadsheet to Make Bar Charts and Pie Charts

To make a graph using a spreadsheet, place the data from the frequency table into the cells. Then select the data, go to the Insert tab, and choose the bar graph or pie chart that you would like. For this example, we will choose a pie graph.



After the spreadsheet has created your pie graph you can choose which design you prefer by clicking on the Chart Design tab. Since these pie pieces represent car colors, we matched the color of each wedge to the color of the car in our pie chart above.

To give your graph a meaningful title, click on Chart Title. There are many other settings that you can experiment with.

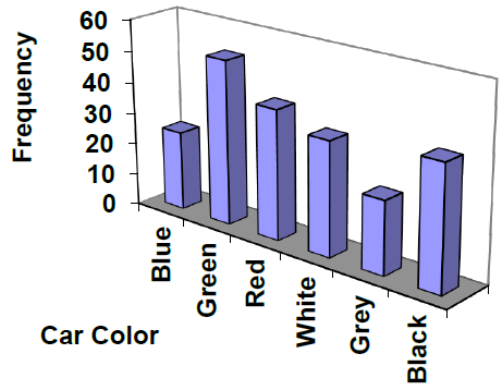


3.2.8 Misleading Graphs

Graphs can be misleading intentionally or unintentionally. It's better to keep them simple, clear and well-labeled. People sometimes add features to graphs that don't help convey their information.

Example 3.2.6

3-dimensional bar chart like the one shown is usually not as effective as a 2-dimensional graph. The extra dimension does not add any useful information.



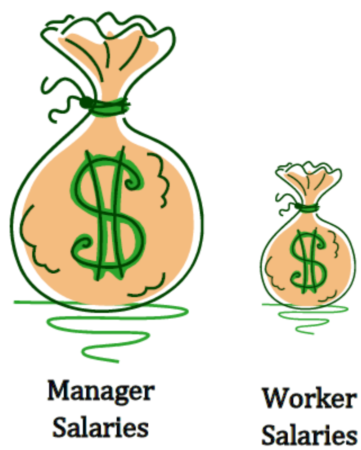
Here is another way that fanciness can sometimes lead to trouble. Instead of plain bars, it is tempting to substitute images. This type of graph is called a pictogram. □

3.2.9 Perceptual Distortion

A ***pictogram*** is a statistical graphic in which the size of the picture is intended to represent the frequency or size of the values being represented. We need to be careful with these, because our brains perceive the relationship between the areas, not the heights.

Example 3.2.7 A labor union might produce this graph to show the difference between the average manager salary and the average worker salary.

The average manager salary is twice as high as the average worker salary as in a bar graph, but the image is also twice as wide. That makes it look like the manager salary is 4 times as large as the worker salary. The area needs to accurately portray the relationship, otherwise we will have a perceptual distortion.

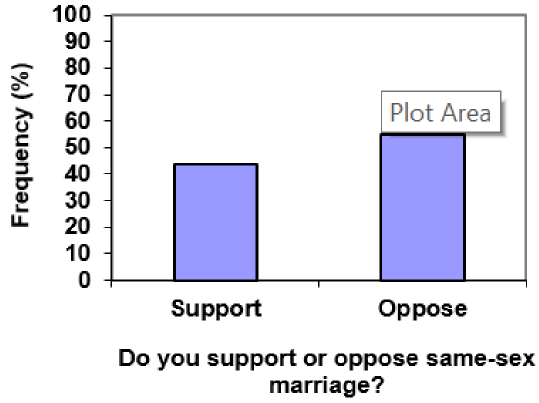


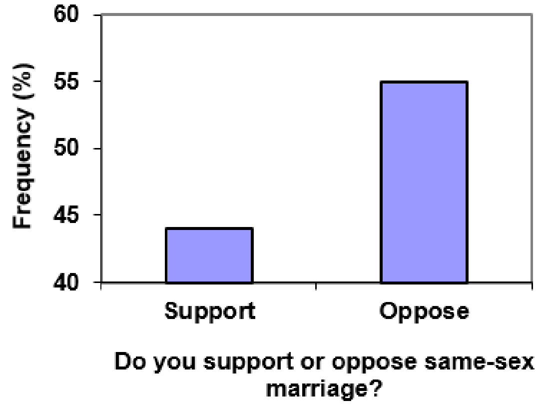
□

3.2.10 Misleading Scale

Another type of 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 3.2.8 Compare the two graphs below showing support for same-sex marriage rights from a poll taken in December, 2008[1]. At a glance, the two graphs suggest very different stories. The second graph makes it look like more than three times as many people oppose marriage rights as support them. But when we look at the scale we can see that the difference is about 12%. By not starting at zero the difference looks enlarged.





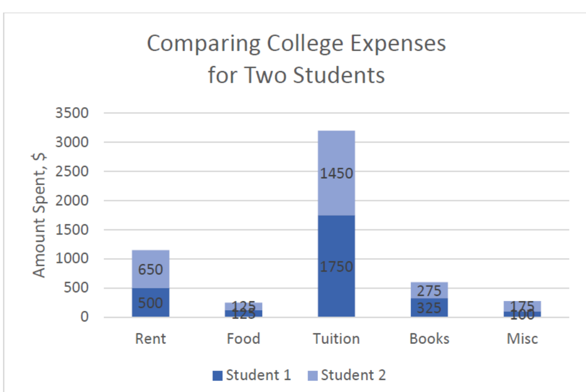
□

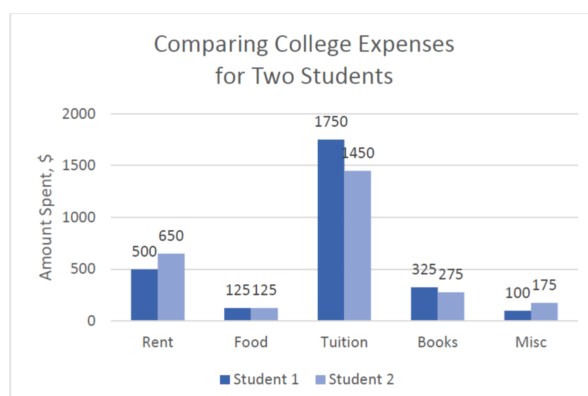
3.2.11 Stacked Bar Graphs

Another type of graph that can be hard to read and sometimes misleading is a stacked bar graph. In a ***stacked bar graph***, the values we are comparing are stacked on top of each other vertically.

Example 3.2.9 The table lists college expenses for two different students and we want to compare them. A stacked bar graph shows the expenses stacked vertically, but we are interested in the differences, not the totals.

Expense	Student 1	Student 2
Rent	\$500	\$650
Food	\$125	\$125
Tuition	\$1750	\$1450
Books	\$325	\$275
Misc	\$100	\$175





It is much easier to interpret the differences in a side-by-side bar chart. □

3.2.12 Presenting Quantitative Data

With categorical data, the horizontal axis is the category, but with quantitative, or numerical, data we have numbers. If we have repeated values we can also make a frequency table.

[1]CNN/Opinion Research Corporation Poll. Dec 19-21, 2008, from http://www.pollingreport.com/civil.htm

Example 3.2.10 A teacher records scores on a 20-point quiz for the 30 students in their 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 and 0.

Here is a frequency table with the scores grouped and put in order.

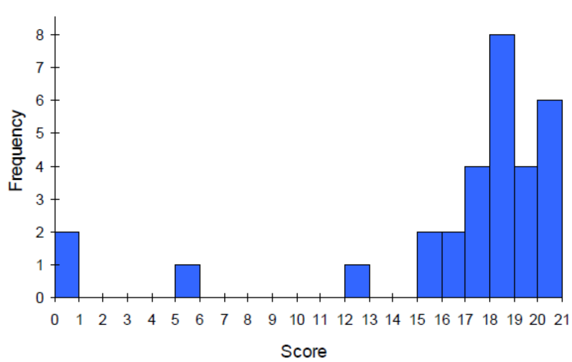
Quiz Score	***Frequency of Students***
0	2
15	1
12	1
15	2
16	2
17	4
18	8
19	4
20	6

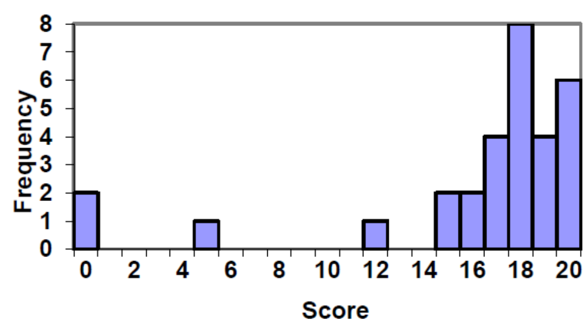
□

Using this table, 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 wouldn't make sense; the first and second bars would be five values apart, while the later bars would only be one value apart. Instead, we will to treat the horizontal axis as a number line. This type of graph is called a histogram.

3.2.13 Histograms

A ***histogram*** is like a bar graph, but the horizontal axis is a number line. Unlike a bar graph, there are no spaces between the bars. Here are two histograms for the data given above. Notice that in the one on the left, the two scores of 15 are to the right of 15, or between 15 and 16. The horizontal scales on histograms can be confusing for this reason. Some people choose to have bars start at ½ values to avoid this ambiguity, as in the one on the right.





If we have a large number of different data values, a frequency table listing every possible value would be way too long. There would be too many bars on the histogram to reveal any patterns. For this reason, it is common with quantitative data to group data into class intervals.

3.2.14 Class Intervals

Class intervals are groupings of the data. In general, we define class intervals so that:

1. Each interval is equal in size. For example, if the first class contains values from 120-134, the second

class should include values from 135-149.

2. We typically have somewhere between 5 and 20 classes, depending on the number of data values we're working with.

In the next example, we'll make a histogram using class intervals.

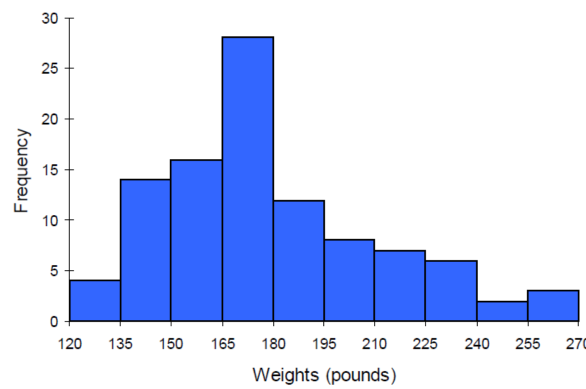
Example 3.2.11 Suppose we have collected weights from 100 subjects who identify as male, as part of a nutrition study. For our weight data, we have values ranging from a low of 121 pounds to a high of 263 pounds, giving a total span of $263 - 121 = 142$.

There are many ways to draw a histogram and we will explain one way as an example. We will create 10 classes by doing the following calculation: Take the range of 142 and divide it by 10 to find the class width.

$142/10 = 14.2$

Then we round up the class width so that the largest data value will be in one of the classes. So we round up to a class width of 15. Since the minimum data value is 121, we will choose 120 to start with since it is a multiple of 15 that is less than 121.

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



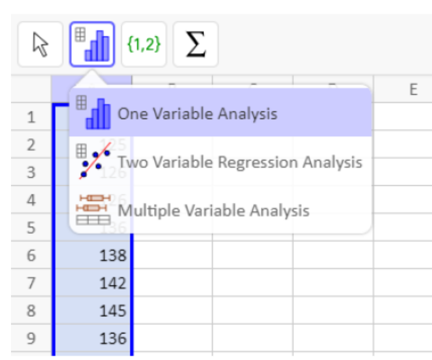
□

When using class intervals, it is much easier to use technology that was specifically designed to make histograms. GeoGebra is one program that lets you adjust the class widths to see which graph best displays the data.

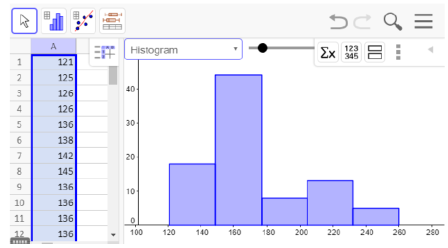
3.2.15 Histograms Using Technology

We will be using GeoGebra throughout this chapter to make graphs and calculate summary statistics. There is an online version and one you can download available at www.Geogebra.org. The instructions are similar for both.

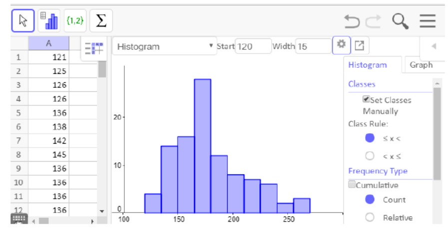
The first thing we need to do is enter the data in GeoGebra's spreadsheet. You can access the spreadsheet from Main Menu →View →Spreadsheet. Next, enter your data and select that column. Then click on the histogram icon in the menu bar on the left side and select One Variable Analysis.



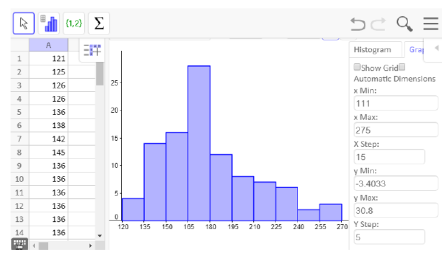
A new window will pop up showing a visual of the data. There is a drop-down menu for the type of graph, but histogram is the default. Notice that the bars are not lined up with the tick-marks at the bottom, so we want to edit this histogram. The slider bar at the top will let you see different class widths, but we want to choose our class widths manually.



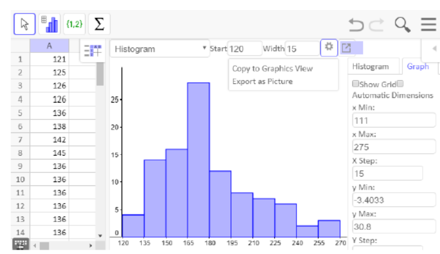
If you close the menu at the top right by clicking on the left pointing triangle, you will see a settings wheel. Click on the wheel and check the box for set classes manually. To match our previous histogram, we will start at 120 pounds and set a class width of 15 pounds.



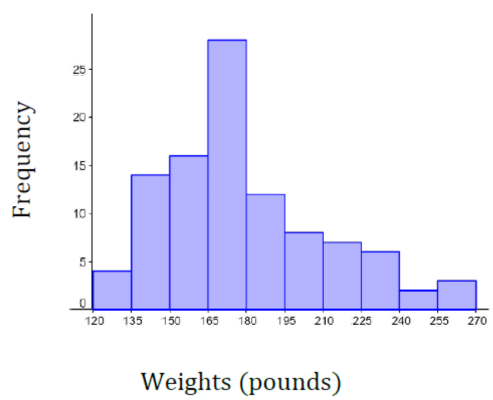
Now the bars of the histogram match our previous graph, but we need to edit the axis labels to match. Click on the graph tab on the right side and uncheck the box for automatic dimensions. We set the x min, x max, x step, y min, y max and y step as shown.



To put the graph in an assignment or a book such as this one, select the export icon and choose Export as Picture. The downloaded version also has a Copy to Clipboard option. Then insert the graph into any document and add axis labels.



Here is our finished histogram:



3.2.16 The Shape of a Distribution

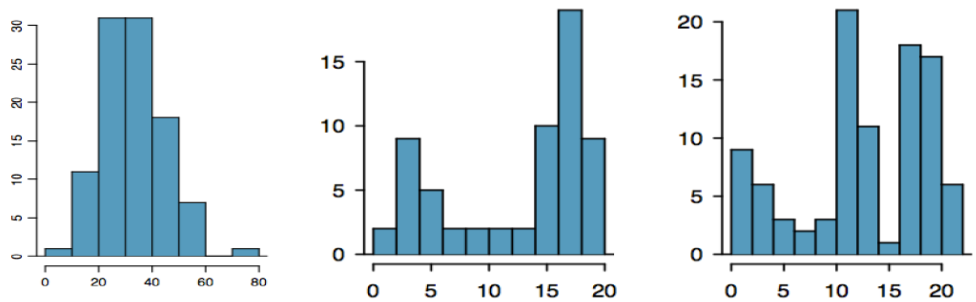
Once we have our histogram, we can use it to determine the shape of the data or ***distribution***. When describing distributions, we are going to look at four characteristics: shape, center, spread and outliers. Center and spread (variation) will be covered in the next two sections.

3.2.17 Modality

The ***modality*** of a distribution indicates the number of peaks or hills in its histogram.

- It is unimodal if it has one peak.
- It is bimodal if it has two peaks.
- It is multimodal if it has multiple peaks.

Example 3.2.12 The first graph is ***unimodal***, the second is ***bimodal*** and the third is ***multimodal***.



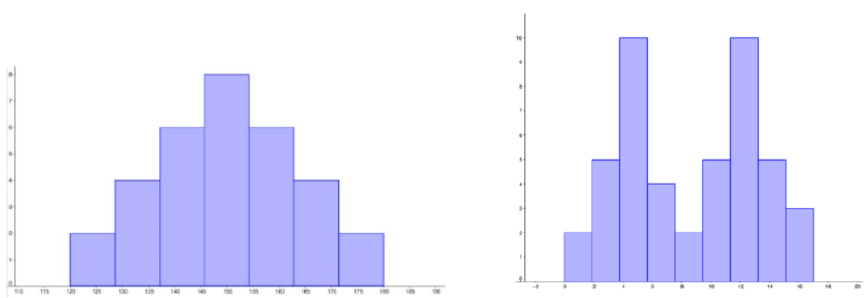
□

A bimodal distribution can result when two different populations have been grouped together and they are overlapping. It would be better to separate them into two separate graphs. For example, the grams of sugar per serving in sugar and non-sugar cereals.

3.2.18 Symmetry

A distribution is ***symmetric*** if the left side of the graph mirrors the right side.

Example 3.2.13 The graph on the left is symmetric and unimodal while the graph on the right is roughly symmetric and bimodal.



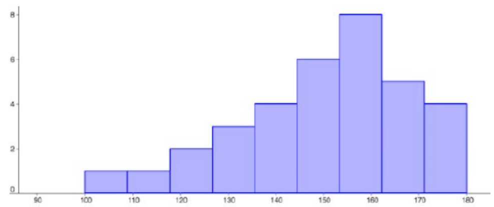
□

3.2.19 Skewness

If a distribution is not symmetric then we say it is skewed. A graph can be ***skewed to the left*** or ***skewed to the right***. We say it is skewed in the direction of the longer tail.

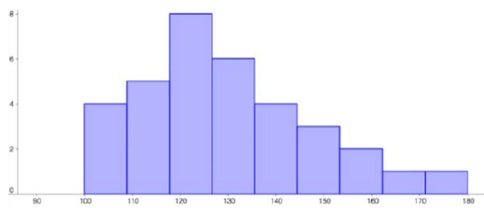
3.2.20 Skewed to the Left

A left skewed graph is also called a ***negatively skewed*** graph. The longer tail will be on the left or negative side.



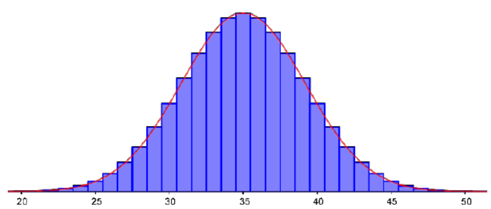
3.2.21 Skewed to the Right

A right skewed graph is also called a ***positively skewed*** graph. The longer tail will be on the right or positive side.



3.2.22 The Normal Distribution

The ***normal distribution*** has a very specific shape. It is unimodal and symmetric with a bell-shaped graph.

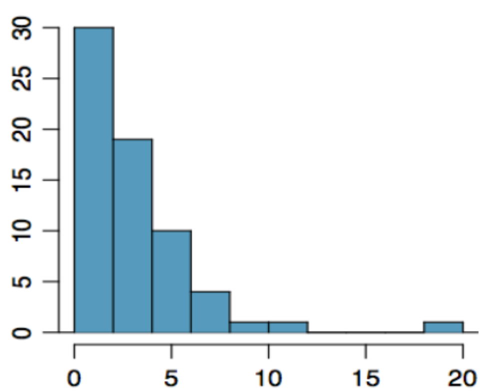


3.2.23 Outlier

Outliers are data values that are unusually far away from the rest of the data. There is often a gap between the outlier and the rest of the graph. This visual determination of outliers is often subjective and depends on the situation.

Example 3.2.14

In the graph to the right we have a unimodal distribution that is skewed to the right. There appears to be an outlier near 20.



□

3.2.24 Exercises

1. True or False: The bars of a histogram should always touch.

2. True or False: The bars of a bar graph should always touch.

3. Is the data described categorial or quantitative?

a. In a study, you ask the subjects their age in years.

b. In a study, you ask the subjects their gender.

c. In a study, you ask the subjects their ethnicity.

d. The daily high temperature of a city over several weeks.

e. A person's annual income.

4. Is the data described categorical or quantitative?

a. In a study you ask the subjects how many siblings they have.

b. In a study you ask the subjects what their favorite movie genre is.

c. In a study to measure the subjects' blood pressure.

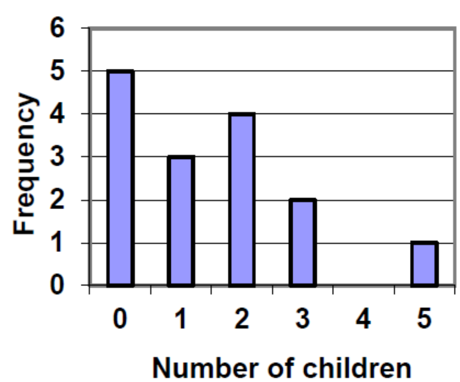
d. The daily rainfall in a city over several weeks.

e. In a study you ask the subjects the amount they spend on housing each month

5. What types of graphs are used for categorical data?

6. What types of graphs are used for quantitative data?

7. A group of adults were asked how many children they have in their family. The bar graph to the right shows the number of adults who indicated each number of children.



a. How many adults had 3 children?

b. How many adults where questioned?

c. What percentage of the adults questioned had 0 children?

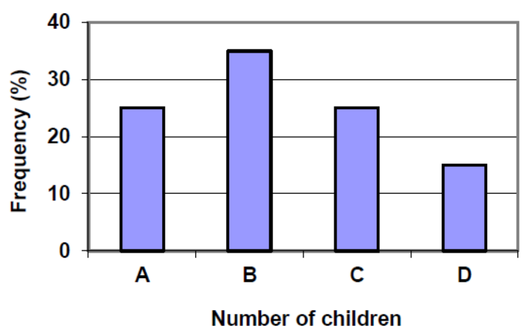
8. Jasmine was interested in how many days it would take a DVD order from Netflix to arrive at her door. The graph shows the data she collected.

a. How many movies took 2 days to arrive?

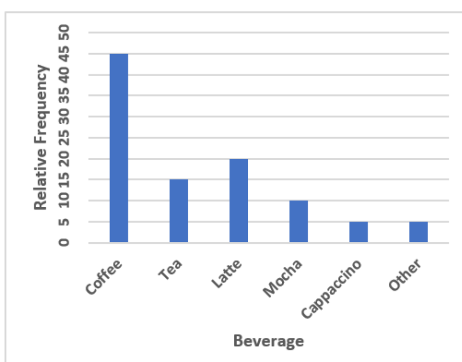
b. How many movies did she order in total?

c. What percentage of the movies arrived in one day?

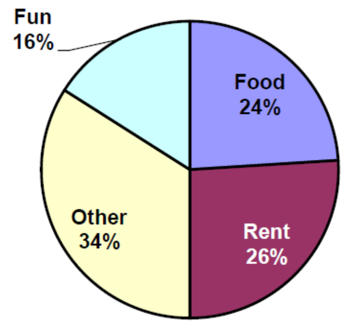
9. This relative frequency bar graph shows the percentage of students who received each letter grade on their last English paper. The class contains 20 students. What number of students earned an A on their paper?



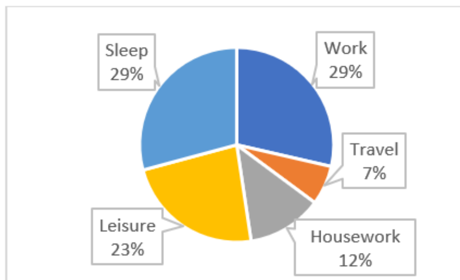
10. This relative frequency bar graph shows the percentage of each drink type served over the weekend at a local coffee shop. There were 120 drinks served in total. How many served drinks were lattes?



11. Corey categorized his spending for this month into four categories: Rent, Food, Fun, and Other. The percentages he spent in each category are pictured here. If he spent a total of $2,600 this month, how much did he spend on rent?



12. Habiba categorized the amount of time spent each week into 5 categories: Work, Travel, Housework, Leisure, and Sleep. If there are a total of 168 hours each week, how many hours does Habiba spend travelling each week?



13. In a survey[2], 1012 adults were asked whether they personally worried about a variety of environmental concerns. The number of people who indicated that they worried "a great deal" about some selected concerns is listed below.

a. Is this categorical or quantitative data?

b. Make a bar chart for this data.

c. Why can't we make a pie chart for this data?

Environmenal Issue	Frequency
Pollution of drinking water	597
Contamination of soil and water by toxic waste	526
Air pollution	455
Global warming	354

14. In a survey, 2056 adults were asked about their views on immigration. The percent of people who responded that immigrants to the United States are making each of the following situations in the country better are listed below.

a. Is this categorical or quantitative data?

b. Make a relative frequency bar chart for this data.

c. Can we make a pie chart for this data?

[2]Gallup Poll. March 5-8, 2009. http://www.pollingreport.com/enviro.htm

Situation	Relative Frequency (%)
Food, music and the arts	57
The economy in general	43
Social and moral values	31
Job opportunities for you and your family	19
Taxes	20
Crime	7

15. The following table is from a sample of five hundred homes in Oregon that were asked the primary source of heating in their home.

a. How many of the households heat their home with firewood?

b. What percent of households heat their home with natural gas?

Type of Heat	Relative Frequency (%)
Electricity	33
Heating Oil	4
Natural Gas	50
Firewood	8
Other	5

16. The following table is from a sample of 50 undergraduate students at Portland State University.

a. What percent of the sampled students are below senior class?

b. How many of the sampled students are freshmen?

Class	Relative Frequency (%)
Freshman	18
Sophmore	13
Junior	23
Senior	46

17. A group of adults were asked how many cars they had in their household.

a. Is this categorical or quantitative data?

b. Make a relative frequency table for the data.

c. Make a bar chart for the data.

d. Make a pie chart for 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

18. The table below shows scores on a math test.

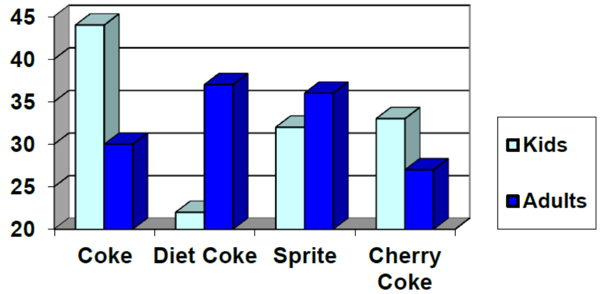
a. Is this categorical or quantitative data?

b. Make a relative frequency table for the data using a class width of 10.

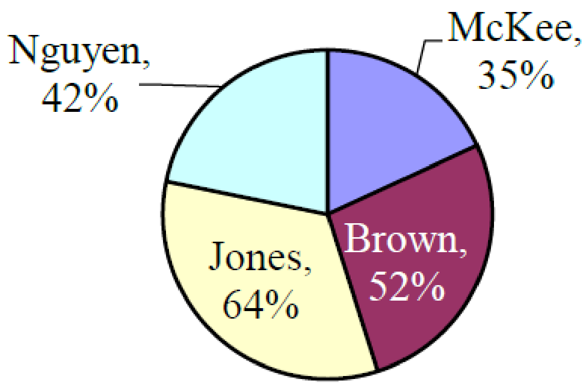
c. Construct a histogram of the data.

82	55	51	97	73	79	100	60	71	85	78	59
90	100	88	72	46	82	89	70	100	68	61	52

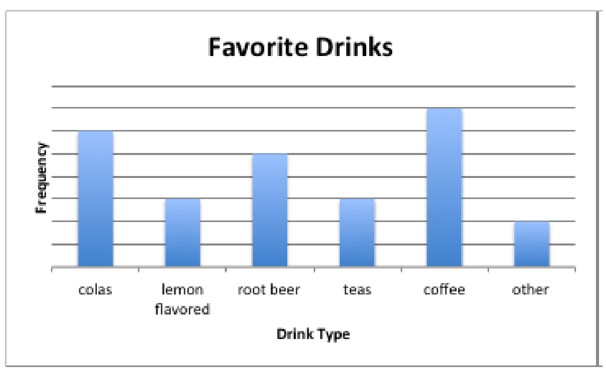
19. This graph shows the number of adults and kids who prefer each type of soda. There were 130 adults and kids surveyed. Discuss some ways in which the graph could be improved.



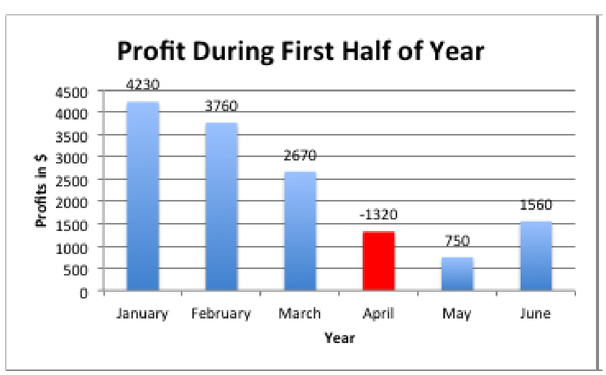
20. A poll was taken asking people if they agreed with the positions of the 4 candidates for a county office. Does this pie chart present a good representation of this data? Explain.



21. Why is this a misleading or poor graph?



22. Why is this a misleading or poor graph?



23. Match each description to one of the graphs.

a. Normal distribution

b. Positive or right skewed

c. Negative or left skewed

d. Bimodal

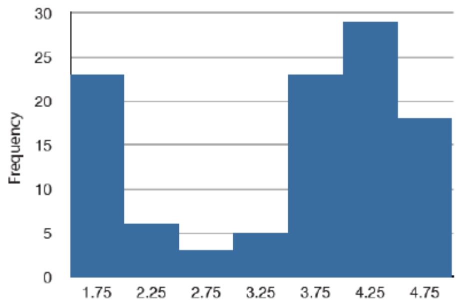


Figure 3.2.15 The frequency of times between eruptions of the Old Faithful geyser.

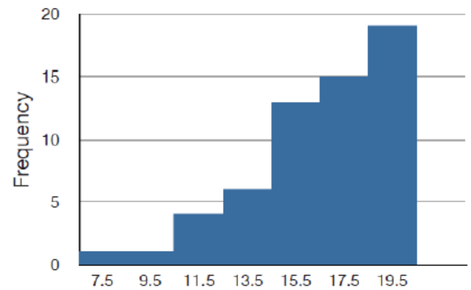


Figure 3.2.16 Scores on a 20-point statistics quiz.

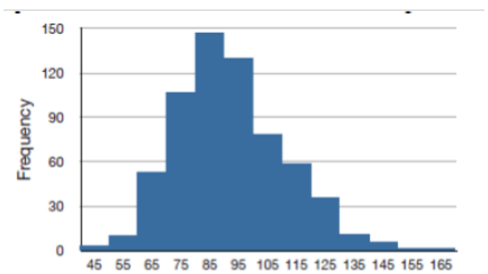


Figure 3.2.17 The distribution of scores on a psychology test.

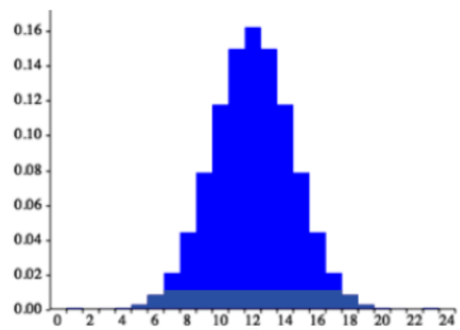


Figure 3.2.18 The number of heads in 24 sets of 100 coint flips.

24. Write a sentence or two to describe each distribution in terms of modality, symmetry, skewness and outliers.

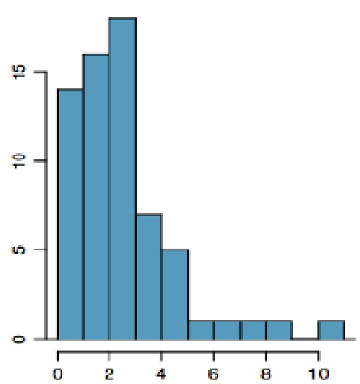


Figure 3.2.19

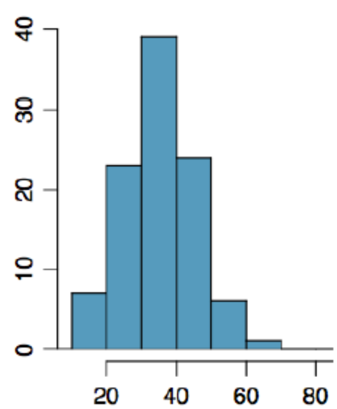


Figure 3.2.20

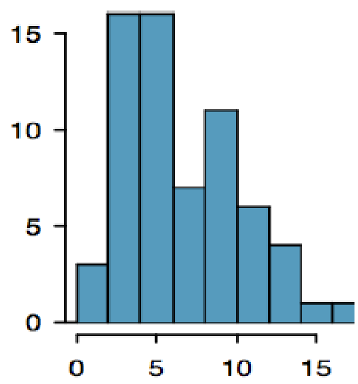


Figure 3.2.21

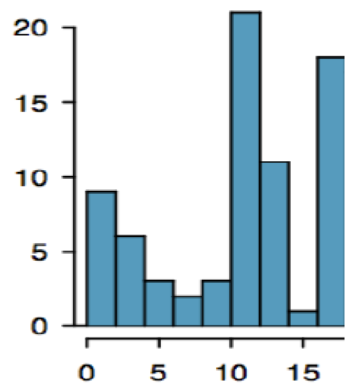


Figure 3.2.22

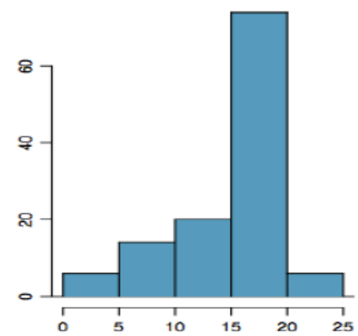


Figure 3.2.23

25. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment. Suppose that a new cancer drug is currently under study. 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 cancer patients throughout their treatment. The following data (in months) are collected.

a. Create a histogram for each dataset, using the same class intervals and scales so you can compare them.

b. Compare and contrast the two distributions.

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, 34

Researcher 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

3.3 Summary Statistics: Measures of Center

Objectives: Section 3.3 Summary Statistics: Measures of Center

Students will be able to:

- Calculate and describe the measures of center: mean and median
- Analyze the relationship of the mean and median to the shape of the data

3.3.1 Calculating Summary Statistics

In addition to graphical and verbal descriptions, we can use numbers to summarize quantitative distributions. We want to know what an "average" value is (where the data is centered), and how spread out the values are. Together, the center and spread provide important information which can be used to estimate our population parameters. In this section we will discuss the measures of center and in the next section we will discuss the measures of spread.

3.3.2 Measures of Center

There are a few different types of "averages" that measure the center, and the one we use will depend on the shape of the distribution. We will mention the mode but focus mainly on the two most common "averages": the ***mean*** and the ***median***.

3.3.3 Mode

In the previous section, we saw that the modes are related to the peaks where similar values are grouped. The mode is the value with the highest frequency (A mode is the value where a peak occurs. One additional way to calculate the mode(s) is to take the midpoint of each peak in the histogram.)

3.3.4 Mean

The ***mean***, or more formally the arithmetic mean, is what probably comes to mind when you hear the word average. The calculation of the mean uses every data value in the distribution and is therefore strongly affected by skew and outliers.

To calculate the mean of a distribution, we divide the sum of the data values by the number of data values we have. The sample mean is usually represented by, a lower-case with a bar over it, read-bar. The lower-case letter n is used to represent the number of data values or ***sample size***.

Mean.

$$\bar{x} = \frac{\text{sum of data values}}{n}$$

Example 3.3.1 Mirabel's exam scores for her last math class were: 79, 86, 82, 94. What is her mean test score?

Solution. To find the mean test score we need to find the sum of her test scores, then divide the sum by the number of test scores ($n = 4$). The mean is:

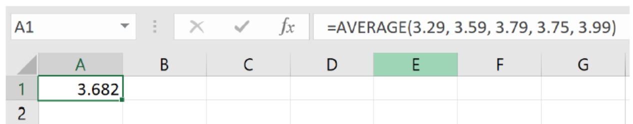
$\bar{x} = \frac{79+86+82+94}{4} = 85.25$ points

We will round the sample mean to one more decimal place than the original data. In this case, we would round 85.25 to 85.3 points. Also notice that the mean has the same units as the data and it is important to label it. □

It is reasonable to calculate the mean by hand when the data set is small, but if the data set is large, or if you will be finding additional statistics, then technology is the way to go. We can find the mean of a data set using the spreadsheet formula `=AVERAGE`.

Example 3.3.2 The price of peanut butter at 5 stores was \$3.29, \$3.59, \$3.79, \$3.75, and \$3.99. Find the mean price using a spreadsheet. There are two ways to use the `=AVERAGE` formula. If your data set is not too large, you can enter each value directly into the formula. Using this method, we write

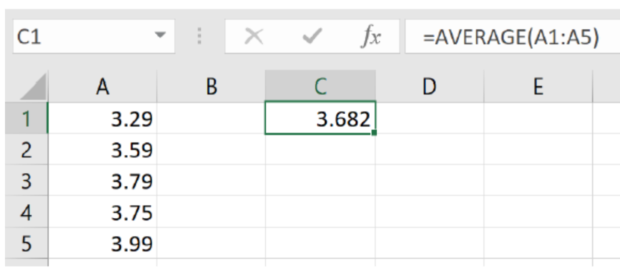
```
=AVERAGE(3.29 3.59, 3.79, 3.75, 3.99)
```



and we get an answer of \$3.68.

The other method is to enter the data values into a single column (or row) of the spreadsheet and reference the column (or row) range in the formula. We can enter the range by highlighting the data values. As illustrated below, if we enter the data into column A, the formula is

```
=AVERAGE(A1:A5)
```



and we also get an answer of \$3.68. □

Sometimes when there is a lot of data with repeated values we are given a frequency table.

Example 3.3.3 One hundred families from a particular neighborhood are randomly selected and asked to give their annual household income rounded to the nearest \$5,000. The results are shown in the frequency table below.

Income (thousands of dollars)	Frequency
\$15	6
\$20	8
\$25	11
\$30	17
\$35	19
\$40	20
\$45	12
\$50	7

Calculating the mean by hand could get tedious if we try to type in all 100 values:

$$\bar{x} = \frac{\overbrace{15 + \cdot + 15}^{6 \text{ terms}} + \overbrace{20 + \cdot + 20}^{8 \text{ terms}} + \overbrace{25 + \cdot + 25}^{11 \text{ terms}} + \cdot}{100}$$

We could calculate this more easily by noticing that adding 15 to itself six times is the same as $(15)(6) = 90$. Using this simplification, we get

$$\begin{aligned}
\bar{x} &= \frac{(15)(6) + (20)(8) + (25)(11) + (30)(17) + (35)(19) + (40)(20) + (45)(21) + (50)(7)}{100} \\
&= \frac{3390}{100} \\
&= 33.9
\end{aligned}$$

The mean household income of our sample is 33.9 thousand dollars or $33,900. ☐

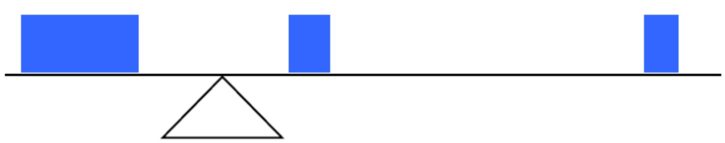
We could also use `=AVERAGE` to find the mean for this example, but it would require entering each repeated value individually. If the mean is all we need, then taking advantage of multiplication as repeated addition is the more straightforward way to go. We could also enter the frequency table and the calculation above in a spreadsheet.

Example 3.3.4 Extending the last example, suppose a new family moves into the neighborhood and has a household income of $5 million ($5000 thousand). Adding this to our sample, our mean becomes:

$$\begin{aligned}
\bar{x} &= \frac{(15)(6) + (20)(8) + (25)(11) + (30)(17) + (35)(19) + (40)(20) + (45)(12) + (50)(7) + (5000)(1)}{100} \\
&= \frac{8390}{101} \\
&= 83.069
\end{aligned}$$

While 83.1 thousand dollars, or $83,100 is the correct mean household income for the new sample, it is no longer representative of the neighborhood – in fact, it is greater than every income in the sample aside from the new one we added! ☐

Imagine the data values on a see-saw or balance scale. The mean is the value at the tip of the triangle that keeps the data in balance, like in the picture below.



If we graph our household data, the $5 million value is so far out to the right that the mean has to adjust to keep things in balance.

For this reason, when working with data that is skewed or has outliers, it is common to use a different measure of center, the median.

3.3.5 Median

The ***median*** of a data set is the “middle” value, when the data are listed in order from smallest to largest. We can also think of the median as the value that has 50% of the data below it and 50% of data above it. As we will discover later, the median is also the ***50th percentile***.

Median.

If the number of data values is odd, then the median is the middle data value If the number of data values is even, then the median is the mean of the middle pair

Example 3.3.5 When finding the median of an odd number of values.

Use the following quiz scores: 5, 10, 8, 6, 4, 8, 2, 5, 7, 7, 6

We must start by listing the data in order: 2, 4, 5, 5, 6, 6, 7, 7, 8, 8, 10.

It is helpful to mark or cross off the numbers as you list them to make sure you don’t miss any. Also, be sure to count the number of data values in your ordered list to make sure it matches the number of data values in the original list.

In this example there are 11 quiz scores. When the distribution contains an odd number of data values there will be a single number in the middle and that is the median. For small data sets, we can “walk” one value at a time from the ends of the ordered list towards the center to find the median

$$\overbrace{2,4,5,5,6}^{\text{Lower Half}}\quad\overbrace{6}^{\text{Median}}\quad\overbrace{7,7,8,8,10}^{\text{Upper Half}}$$

The median test score is 6 points. □

Example 3.3.6 When finding the median of an even number of values.

Use the following quiz scores: 2, 4, 5, 5, 6, 6, 7, 7, 8, 8, 10, 20.

There are now 12 quiz scores in our sample. When the distribution contains an even number of data values there will be a pair of values in the middle rather than a single value. Then we take the average of the middle two values.

$$\overbrace{2,4,5,5,6}^{\text{Lower Half}}\quad\overbrace{6,7}^{\text{Middle Pair}}\quad\overbrace{7,8,8,10,20}^{\text{Upper Half}}$$

Median= $\frac{6+7}{2} = 6.5$ points

What is important to notice is that despite adding an outlier to our data set, the median is largely unaffected. The median quiz score for the new distribution is 6.5 points.

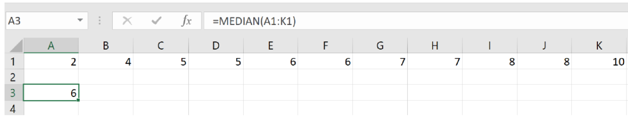
We can also find the median using the spreadsheet formula `=MEDIAN`. Just like the spreadsheet function `=AVERAGE`, we can either list the individual data values in the formula, or we can enter the data values into a row (or column) and use the row range (or column range) in the formula.

Using the data values of the original distribution, we can write function as

```
=MEDIAN(2, 4, 5, 6, 6, 7, 7, 8, 8, 10)
```

or

```
=MEDIAN(A1:AK)
```



and we get 6 points for the median test score. □

Example 3.3.7 Let's continue with our peanut butter example and find the median both by hand and with a spreadsheet. The price of peanut butter at 5 stores was \$3.29, \$3.59, \$3.79, \$3.75, and \$3.99.

To find the median by hand, we must first list the prices in order. This give us: \$3.29, \$3.59, \$3.75, \$3.79, \$3.99

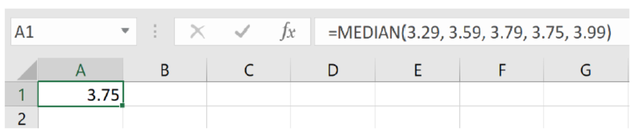
Since there are an odd number of data values in the sample ($n = 5$), we know that the median will be the single data value in the middle of the ordered list.

$$\overbrace{3.29, 3.59}^{\text{Lower Half}}\ \overbrace{3.75}^{\text{Median}}\ \overbrace{3.79, 3.99}^{\text{Upper Half}}$$

The median price of peanut butter at these five stores is \$3.75.

Using a spreadsheet, we write

```
=MEDIAN(3.29, 3.59, 3.79, 3.75, 3.99)
```



and we get a median price of \$3.75. □

It is worth noting that when you use a spreadsheet to find the median you do not have to order the data first. You can enter the data values in the order they are given to you.

3.3.6 The Relationship Between the Mean and the Median

If a distribution is skewed, the mean is pulled in the direction of the skew, as we saw in the see-saw diagram. In a right skewed distribution, the mean is greater than the median, while in a left skewed distribution, the mean is less than the median. If the distribution is symmetric, the mean and the median will be approximately equal.

To demonstrate this, we have entered some data in GeoGebra, as previously explained, and made histograms. To see the statistics that GeoGebra calculates, we click on the summation symbol ($\sum x$) on the right-hand menu bar.

Example 3.3.8 Fifty people from the Portland Metro area who are employed full time were sampled and their annual salaries were recorded (to the nearest thousand dollars). The histogram and summary statistics from GeoGebra are shown below.

From the histogram we can see that the shape of the distribution is unimodal and skewed to the right. We can see from the statistics output on the left that the mean is greater than the median. This is because the few people with higher incomes bring the average up.

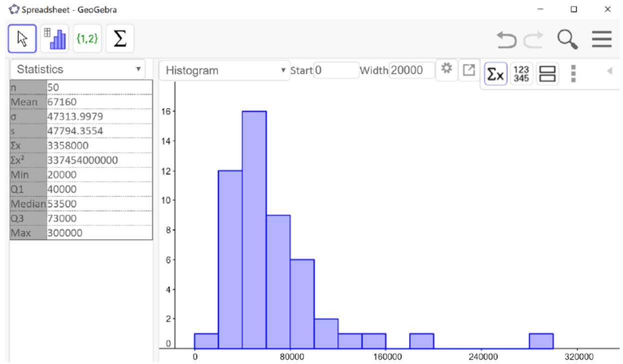


Figure 3.3.9 Mean $= \$67,160$, Median $= \$53,500$, Mean $>$ Median

□

Example 3.3.10 A random selection of 30 math 105 exams at PCC were sampled and their scores were recorded. The histogram of the resulting distribution is shown below.

The shape of the distribution is unimodal and skewed to the left. There also appears to be an outlier between 20 and 30. We can see from the statistics output that the mean is less than the median. This is because the low test score brought the average down. points points

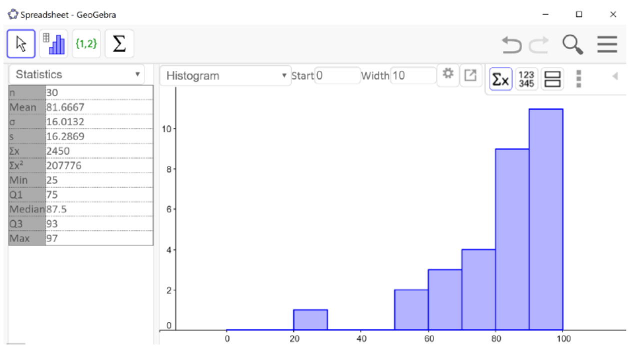


Figure 3.3.11 Mean $= 81.7$ points, Median $= 87.5$ points, Mean $<$ Median

□

Example 3.3.12 Nineteen people identifying as female were sampled and their heights (in inches) were recorded. The histogram of the resulting distribution is shown below.

The shape of the distribution is unimodal and roughly symmetric. We can also see from the statistics output that the mean and the median are approximately equal. inches inches

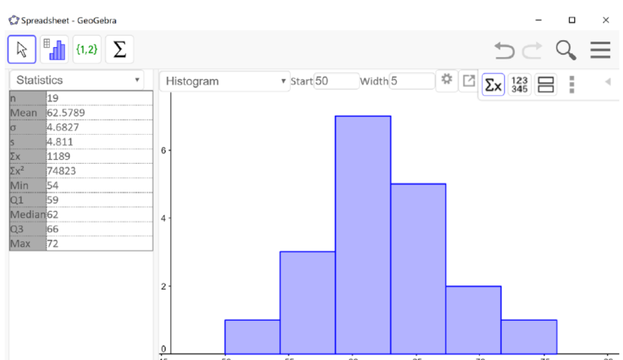


Figure 3.3.13 Mean $= 62.6$ inches, Median $= 62$ inches Mean $\approx$ Median

□

We can use these observations in reverse as well. If we know the mean is greater than the median, then we can expect the distribution to be skewed to the right. If the mean is less than the median, then we can expect the distribution to be skewed to the left. When the mean and the median are approximately equal, the distribution is likely to be symmetric.

Example 3.3.14 Recent college graduates were asked how much student loan debt they have. The data has a mean of $46,265 and a median of $33,652. Just based on this information, do you expect the distribution to be symmetric, skewed to the left, or skewed to the right?

Solution. Since the mean is greater than the median, we can expect the distribution to be skewed to the right. □

3.3.7 Exercises

1. A group of diners were asked how much they would pay for a meal. Their responses were: $7.50, $25.00, $10.00, $10.00, $7.50, $8.25, $9.00, $5.00, $15.00, $8.00, $7.25, $7.50, $8.00, $7.00. $12.00.

a. Find the mean, including units.

b. Find the median, including units.

c. Based on the mean and the median, would you expect the distribution to be symmetric, skewed left, or skewed right? Explain.

2. The amount of commercials in an hour of television varies by channel. The total length (in minutes) of all commercials from 8 pm to 9 pm in for some selected broadcast and cable channels on a weekday evening were: 10, 12.75, 7, 9, 9.75, 6.5, 12.5, 12.5, 8.75, 17, 10.5, 2.

a. Find the mean, inluding units.

b. Find the median, including units.

c. Based on the mean and median, would you expect the distribution to be symmetric, skewed left, or skewed right?

3. 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.

a. Find the mean, including units.

b. Find the median, including units.

c. Based on the mean and the median, would you expect the distribution to be symmetric, skewed left, or skewed right? Explain.

4. You weigh 9 Oreo cookies, and you find the weights (in grams) are: 3.49, 3.51, 3.51, 3.51, 3.52, 3.54, 3.55, 3.58, 3.61.

a. Find the mean, including units.

b. Find the median, including units.

c. Based on the mean and the median, would you expect the distribution to be symmetric, skewed left, or skewed right? Explain.

5. Use the following table is the cost of purchasing a car at a local dealership. Some of the cars sold were new and some were used.

a. Calculate find the mean, including units.

b. Can you figure out how to find the median using the frequency table? See if you can do it without listing out all the data values.

c. Based on the mean and the median, would you expect the distribution to be symmetric skewed left or skewed right? Explain.

Cost (Thousands of dollars)	Frequency
15	3
20	7
25	10
30	15
35	13
40	11
45	9
50	7

6. As part of a study of email, a researcher counted the length of 34 emails. The lengths of the emails are shown below, rounded to the nearest thousand characters (so a length 0 means that the numbers of characters rounded to 0, not that the message was blank).[1]

a. Calculate and find the mean, including units.

b. Can you figure out how to find the median using the frequency table? See if you can do it without listing all the data values.

c. Based on the mean and the median, would you expect the distribution to be symmetric, skewed left, or skewed right?

Length of an email (Thousands of characters)	Frequency
0	4
1	5
2	2
3	3
4	3
5	1
6	3
7	3
8	0
9	3
10	3
11	2
12	0
13	0
14	2

7. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment. Suppose that a new cancer drug is currently under study. 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 cancer patients throughout their treatment. The following data (in months) are collected.

a. Find the mean and median of each group.

b. Compare and contrast the two groups.

- Researcher 1: 3, 4, 11, 15, 16, 17, 22, 44, 37, 16, 14, 24, 25, 15, 26, 27, 33, 29, 35, 44, 13, 21,

[1] The data is from Advanced High School Statistics, 2nd Ed. (https://www.openintro.org/stat/textbook.php?stat_book=aps).

22, 10, 12, 8, 40, 32, 26, 27, 31, 34, 29, 17, 8, 24, 18, 47, 33, 34

- Researcher 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 5.

8. The US Census Bureau, in addition to counting the population of the US every 10 years, conducts yearly informational surveys, such as the American Community Survey (ACS). For the 2012 ACS, a randomly chosen group of 20 respondents (10 males, 10 females) answered a question about their incomes.[2]

- Males: \$53,000; \$70,000; \$12,800; 30,000; \$4,500; \$42,000; \$48,000; \$60,000; \$108,000; \$11,000
- Females: \$1,600; \$1,200; \$20,000; \$25,000; \$670; \$29,000; \$44,000; \$30,000; \$5,800; \$50,000

a. Find the mean and median of each group.

b. Compare and contrast the two groups.

9. An experiment compared the ability of three groups of participants to remember briefly-presented chess positions. The data are shown below. The numbers represent the average number of pieces correctly remembered from three chess positions.

a. Make a histogram for each group.

b. Find the mean of each group.

c. Find the median of each group

d. Compare the shapes of the distributions as well as the centers of the three groups.

Non-players	Beginners	Tournament Players
22.1	32.5	40.1
22.3	37.1	45.6
26.2	39.1	51.2
29.6	40.5	56.4
31.7	45.5	58.1
33.5	51.3	71.1
38.9	52.6	74.9
39.7	55.7	75.9
39.7	55.7	75.9
43.2	55.9	80.3
43.2	57.7	85.3

10. There is evidence that smiling can attenuate judgments of possible wrongdoing. This phenomenon termed the "smile-leniency effect" was the focus of a study by Marianne LaFrance & Marvin Hecht in 1995[3]. The following data are measurements of how lenient the sentences were for three different types of smiles and one neutral control. A higher number indicates greater leniency. The same subject was used for all of the conditions so that may affect the results.

a. Make a histogram for each smile type and the neutral control.

b. Find the mean for each type of smile and the neutral control.

c. Find the median for each type of smile and the neutral control.

d. Compare the shapes of the distributions as well as the centers for each type of smile and control.

[2]Data from Advanced High School Statistics, 2nd Ed. Section 2.1 Exercise #1. (https://www.openintro.org/stat/textbook.php?stat_book=aps).

False Smile	Felt Smile	Miserable Smile	Nuetral Control
2.5	7	5.5	2
5.5	3	4	4
6.5	6	4	4
3.5	4.5	5	3
3	3.5	6	6
3.5	4	3.5	4.5
6	3	3.5	2
5	3	3.5	6
4	3.5	4	3
4.5	4.5	5.5	3
5	7	5.5	4.5
5.5	5	4.5	8
3.5	5	2.5	4
6	7.5	5.5	5
6.5	2.5	4.5	3.5
3	5	3	4.5
8	5.5	3.5	6.5
6.5	5.5	8	3.5
8	5	5	4.5
6	4	7.5	4.5
6	5	8	2.5
3	6.5	4	2.5
7	6.5	5.5	4.5
8	7	6.5	2.5
4	3.5	5	6
3	5	4	6
2.5	3.5	3	2
8	9	5	4
4.5	2.5	4	5.5
5.5	8.5	4	4
7.5	3.5	6	2.5
6	4.5	8	2.5
9	3.5	4.5	3
6.5	4.5	5.5	6.5

11. Make up three data sets with 5 values each that have:

a. The same mean but different medians

b. The same median but different means.

12. The frequency table below shows the number of women's shoes that were sold in an hour at a local shoe store.

a. Would you treat this data as categorical or quantitative?

b. How would the bar graph be different from the histogram?

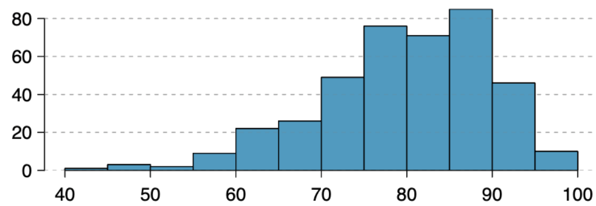
c. Treat the data as quantitative and find the mean and the median. Are these useful statistics?

---

[3]LaFrance, M., & Hecht, M. A. (1995) Why smiles generate leniency. Personality and Social Psychology Bulletin, 21, 207-214. Adapted from www.onlinestatbook.com, by David M. Lane, et al, used under CC-BY-SA 3.0.

Shoe Size	Frequency
5	4
6	4
7	6
8	6
9	5

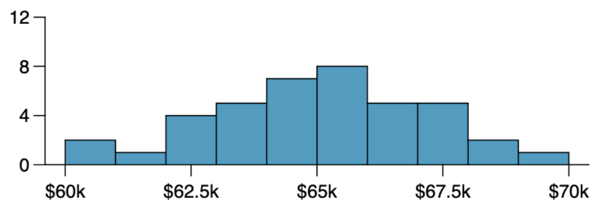
13. At the end of the term, 400 students take a final exam, and their scores (as percentages) are plotted in a histogram.[4]



a. Is the graph shown above skewed right, skewed right, or symmetric? Explain why you chose your answer.

b. Based on your answer to part (a), which would you expect: that the mean is less than the median, the mean is greater than the median, or that the mean and median are equal?

14. The following graph shows the distribution of yearly incomes of 40 patrons at a college coffee shop.[5]



a. Is the graph shown above skewed right, skewed right, or symmetric? Explain why you chose your answer.

b. Based on your answer to part (a), which would you expect: that the mean is less than the median, the mean is greater than the median, or that the mean and median are equal?

15. For each of the following distributions, would you expect that the mean is less than the median, less than the median, or equal to the median? Explain your reasoning.

a. Household incomes in the US

b. Weights of newborn babies

c. The number of children in a household in the US

d. Medical costs for all adults

e. Medical costs for adults in the US older than 65

[4]Graph and description from Advanced High School Statistics, 2nd Ed. Section 2.2 Exercise. (https://www.openintro.org/stat/textbook.php?stat_book=aps)

[5]Graph and description from Advanced High School Statistics, 2nd Ed. Section 2.2 Exercise. (https://www.openintro.org/stat/textbook.php?stat_book=aps)

3.4 Summary Statistics: Measures of Variation

Objectives: Section 3.4 Summary Statistics: Measures of Variation

Students will be able to:

- Calculate and describe the measures of variation: standard deviation, range and interquartile range (IQR)
- Calculate the 5-number summary and construct boxplots by hand and/or using technology (boxplots using technology may be modified or not)
- Compare distributions with side-by-side boxplots and percentiles
- Calculate and apply Z-scores

3.4.1 Measures of Variation

Consider these three sets of quiz scores for a 10-point quiz:

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 data sets have a mean of 5 points and median of 5 points, 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. Section C was not as consistent as section A, but not as widely varied as section B.

Thus, in addition to the mean and median, which are measures of center or the "average" value, we also need a measure of how "spread out"" or varied each data set is.

There are several ways to measure the variation of a distribution. In this section we will look at the ***standard deviation***, ***range*** and the ***interquartile range (IQR)***.

3.4.2 Standard Deviation

The ***sample standard deviation, s***, is a measure of variation that tells us how far, on average, the data values deviate, or are different from, the mean. The mean and standard deviation are paired to provide a measure of center and spread for symmetric distributions.

Sample Standard Deviation.

$$s = \sqrt{\frac{\text{Sum of the squared deviations from teh mean}}{n-1}}$$

where n is the sample size, or the number of data values

We will go through the whole process for calculating the standard deviation. Let's say there is another section of quiz scores:

Section D: 0, 5, 5, 5, 5, 5, 5, 5, 5, 10

The mean quiz score, like Sections A, B and C, is 5 points.

The first step in finding the standard deviation is to find the deviation, or difference, of each data value from the mean. We will do this in a table. You could also use a spreadsheet to do these calculations.

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	$5 - 5 = 0$

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 the deviations.

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	$10 - 5 = 5$	$5^2 = 25$

Next, we add the squared deviations and we 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 the division by n is only done if the data set represents a population. When the data set represents a sample (as it almost always does), we divide by n˘1 (in this case,).

We assume Section D represents a sample, so we will divide by 9. Note that our units are now points-squared since we squared all of the deviations. It is much more meaningful to use the units we started with, so to convert back to points we take the square root.

The sample standard deviation for Section D is

$$s = \sqrt{\frac{50}{9}} \approx 2.36 \text{ points.}$$

For comparison, here is the standard deviation for each section listed above:

$S_A = 0$ points

$S_B = 5.27$ points

$S_C = 0.82$ points

$S_D = 2.36$ points

For the standard deviation, we usually use two more decimal places than the original data. This tells us that on average, scores were 2.36 points away from the mean of 5 points. In summary, here are the steps to calculate the standard deviation by hand.

Calculating the Sample Standard Deviation.

1. Find the deviations by subtracting the mean from each data value
2. Square each deviation
3. Add the squared deviations
4. Compute the square root of the sum by dividing by $n - 1$:

$$s = \sqrt{\frac{\text{Sum of the squared deviations from the mean}}{n-1}}$$

There are a few important characteristics we want to keep in mind when finding and interpreting the standard deviation.

- The standard deviation is never negative. It will be zero if all the data values are equal and get larger as the data spreads out.
- The standard deviation has the same units as the original data and it is important to label it.
- The standard deviation, like the mean, can be highly influenced by outliers.

Example 3.4.1 To continue our peanut butter example, we will find the standard deviation of this sample: \$3.29, \$3.59, \$3.79, \$3.75, and \$3.99.

The first thing we need to find is the sample mean, and we know it is \$3.68 from our previous work. Next, we need to find the deviation from the mean for each data value and square it.

Data Value	***Deviation***	***Deviation Squared***
\$3.29	$3.59 - 3.68 = -0.09$	$(-0.39)^2 = 1521$
\$3.59	$3.59 - 3.68 = -0.09$	$(-0.09)^2 = 0.0081$
\$3.79	$3.79 - 3.68 = 0.11$	$(0.11)^2 = 0.0121$
\$3.75	$3.75 - 3.68 = 0.07$	$(0.07)^2 = 0.0049$
\$3.99	$3.99 - 3.68 = 0.31$	$(0.31)^2 = 0.0961$

The sum of the deviations squared is

$$0.1521 + 0.0081 + 0.0121 + 0.0049 + 0.0961 = 0.2733 \text{ dollars-squared.}$$

The sample standard deviation is

$$s = \sqrt{\frac{0.2733}{4}} \approx \$0.2614$$

Since the units are dollars, we will round to two decimal places rather than two more than the data. This gives us a standard deviation of \$0.26. Together with the mean this tells us that on average, the cost of a jar of peanut butter is \$0.26 away from the mean of \$3.68. □

Calculating the standard deviation by hand can be quite a nuisance when we are dealing with a large data set, so we can also use technology. We use the spreadsheet function `=STDEV.S` to find the *sample* standard deviation. Notice that this is different from the population standard deviation, which uses the function `=STDEV.P`.

Just like the spreadsheet functions `=AVERAGE` and `=MEDIAN`, we can either list the individual data values in the formula, or we can enter the data values into a row or column and use the row or column range in the formula.

Example 3.4.2 The total cost of textbooks for the term was collected from 36 students. Use a spreadsheet to find the mean, median, and standard deviation of the sample.

\$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

Since we are finding more than one statistic for this data set, it is much more efficient to enter the data values into a row or column and reference the range in each of the formulas. We enter the data into column A.

Solution.

For the mean we enter:

`=AVERAGE(A1:A36)`

and get a result of \$299.58.

For the median we enter:

`=MEDIAN(A1:A36)`

and get a result of \$307.50.

For the standard deviation we enter:

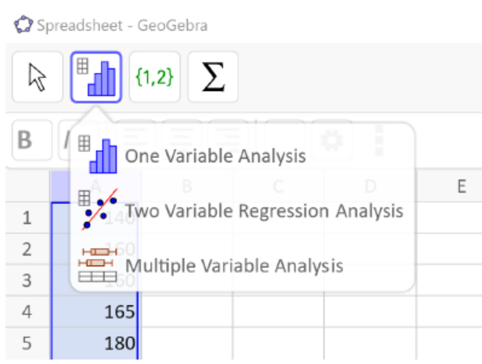
`=STDEV.S(A1:A36)`

and get \$78.68.

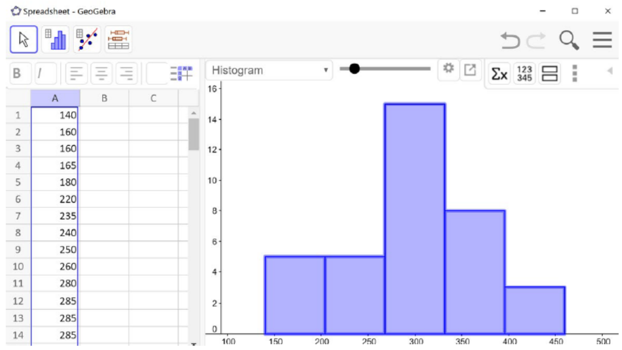
The mean and the median are relatively close to each other, so we can expect the distribution to be approximately symmetric with maybe a slight skew to the left, since the mean is smaller. The mean and standard deviation together tell us that the average cost of textbooks for a term is about \$299.58, give or take \$78.68. □

In addition to a spreadsheet, we will continue our use of GeoGebra. Let's take a look at how to use GeoGeogebra to find the mean, median and standard deviation for the last example. We begin just like we did for making a histogram.

Example 3.4.3 Example 3.4.2 continued: We enter the textbook data into column A of the spreadsheet in GeoGebra. (Main Menu →View →Spreadsheet). Next, select the column title of your data, click on the histogram in the menu bar on the left, and select One Variable Analysis.

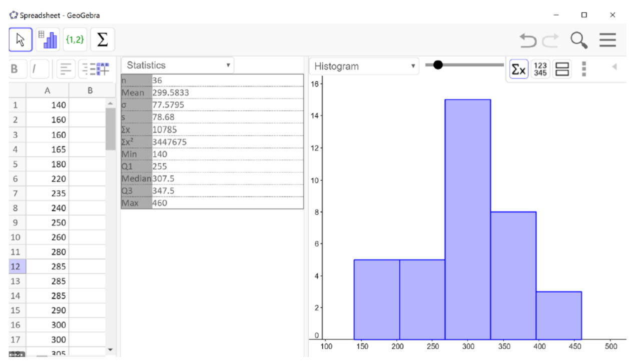


Then you will see the histogram. It is always a good idea to check the shape of your distribution before calculating anything. Notice our histogram matches what we found about the distribution from the mean and median. It is approximately symmetric or slightly skewed to the left.



Next, we click the summation symbol ($\sum x$) in the menu bar on the right. The list of summary statistics will pop up as you can see in the image below. We see that the mean is \$299.58, the median is \$307.50 and the sample standard deviation is \$78.68 – just like we found using the spreadsheet formulas.

The statistics we will use are the sample size (n), the mean, median, and the sample standard deviation (s). The last five entries in the table – min, Q1, median, Q3, and max – together make up the 5-number summary which we will learn about shortly!



The standard deviation is the measure of variation that we pair with the mean for approximately symmetric distributions. This pairing should make sense because the standard deviation uses the mean in its calculation. But what about the median? What measure of variation do we pair with it? □

3.4.3 Range

One candidate is the ***range***. The range tells us the spread or width of the entire data set. We calculate the range as the difference between the maximum and minimum value.

Range.

Range = Max - Min

However, the range is not a very good measure of variation since it is very strongly affected by skew and outliers. Consider, for example, the distribution of full time salaries in the United States. Many people earn a minimum wage salary, while others like Jeff Bezos (Amazon) and Bill Gates (Microsoft) earn millions (if not billions!). A range this large does very little to help us get a sense of the spread where most of the data values lie.

3.4.4 Quartiles and the Interquartile Range

Instead, the measure of variation that we pair with the median is the ***interquartile range (IQR)***. The IQR tells us the width of the middle 50% of data values. By cutting off the lower and upper 25% of data values, we are able to ignore extreme values and provide a more accurate sense of how spread out the distribution is.

The IQR is calculated as the difference between the third quartile (Q3) and the first quartile (Q1). Before we can calculate the interquartile range, though, we need to learn how to find the first and third quartiles.

Interquartile Range (IQR).

$$\text{IQR} = Q_3 - Q_1$$

As the name implies, ***quartiles*** are values that divide the data into quarters. The ***first quartile (Q1)*** is the value that 25% of the data lie below. The ***third quartile (Q3)*** is the value that 75% of the data lie below. As you might have guessed, the second quartile is the same as the median since 50% of the data values lie below it.

We have seen that the data is split in half by the median so to split it into quarters, we find the median of each half of the data.

Quartiles.

Q_1 the median of the lower half of the data

Q_2 is the median of the whole data set

Q_3 is the median of the upper half of the data

If there is an odd number of data values, we don't use the median in either half

Example 3.4.4 (even number of data values): Suppose we have measured the height, in inches, of 12 people who identify as female. The data values are listed below. Find the interquartile range.

59 60 69 64 70 72 66 64 67 66 63 61

Solution. Just like when finding the median, we must first order the data.

59 60 61 63 64 64 66 66 67 69 70 72

Then we divide the data into two halves. In the case of an even sample size, we split the distribution down the middle. The first 6 data values are the lower half and the next 6 data values are the upper half. Then we find the median of each. The median of the lower half is Q1 and the median of the upper half is Q3.

$$\overbrace{59, 60, \overset{Q1=\frac{61+63}{2}=62}{\underbrace{61, 63,}} \; 64, 64,}^{\text{Lower Half}}\overbrace{66, 66, \overset{Q3=\frac{67+69}{2}=68}{\underbrace{67, 69,}} \; 70, 72}^{\text{Upper Half}}$$

In this data $Q_1 = 62$ inches, and $Q_3 = 68$ inches. Then we subtract to find the IQR.

$$IQR = Q_3 - Q_1 = 68 - 62 = 6 \text{ inches}$$

This tells us the that the middle 50% of the women's heights lie within an interval of 6 inches. □

Example 3.4.5 (odd number of data values): Suppose we added one more height (68 inches) to the data set from the previous example. We will again find the interquartile range of the heights.

59 60 61 63 64 64 66 66 67 68 69 70 72

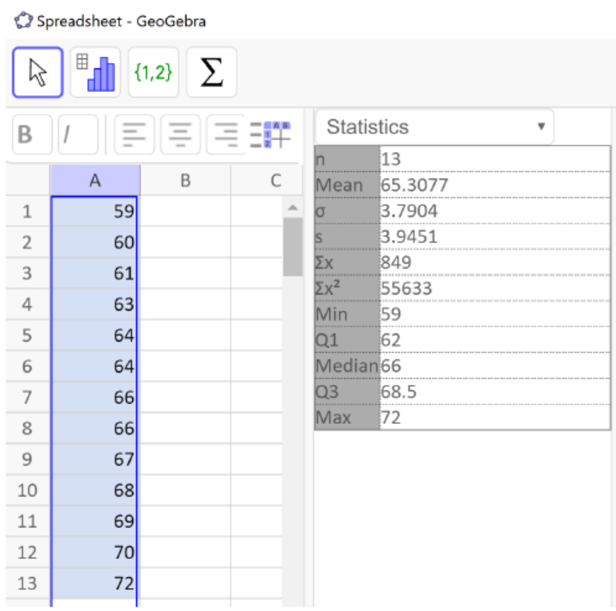
The data are already in order, but we have an odd number of values. To deal with this we do not use the median in the upper or lower halves. The lower half will include the values strictly below the median, and the upper half will include the values strictly above the median.

$$\overbrace{59, 60, \quad \underbrace{\overset{Q_1=\frac{61+63}{2}=62}{61, 63},} \quad 64, 64,}^{\text{Lower Half}} \quad \overbrace{66}^{\text{Median(Ignore)}} \quad \overbrace{66, 67, \quad \underbrace{\overset{Q_3=\frac{68+69}{2}=68.5}{68, 69},} \quad 70, 72}^{\text{Upper Half}}$$

Then the interquartile range is:

$$IQR = Q_3 - Q_1 = 68.5 - 62 = 6.5 \text{ inches}$$

If the data set is small, we can find the first and third quartiles by hand, but we have also seen that they are part of the output from GeoGebra. Here is the output for this data.



From the list of summary statistics we can see that $= Q_1 = 62$ and $= Q_3 = 68.5$ inches. Now we can calculate the interquartile range.

$$IQR = 68.5 - 62 = 6.5 \text{ inches}$$

□

This is the same value we found by hand. It is important to note that we are not using spreadsheets for the five-number summary because they do not calculate the quartiles in the same way, so they will not give the same results. Now that we have learned how to find the quartiles we can make a five-number summary and boxplot.

3.4.5 The Five-Number Summary and Boxplots

The ***five-number summary*** is made up of the minimum, Q1, median, Q3, and the maximum. These five values divide the data into quarters. A ***boxplot***, also called a ***box-and-whisker plot***, is a graphical representation of the five-number summary. Each region of the boxplot contains approximately the same number of data values, so we can see the spread for each region. We can find the five-number summary and draw a boxplot by hand or by using GeoGebra. In our last example the five-number summary from GeoGebra is: 59, 62, 66, 68.5, 72 inches.

Five-Number Summary.

Minimum, Q_1, Median, Q_3, Maximum

We will use GeoGebra to find the five-number summary for the next example and then explain how to draw a boxplot.

Example 3.4.6 Let's continue with the cost of textbook data from Example 3.4.2. Use GeoGebra to find the five-number summary for this sample and draw a boxplot by hand.

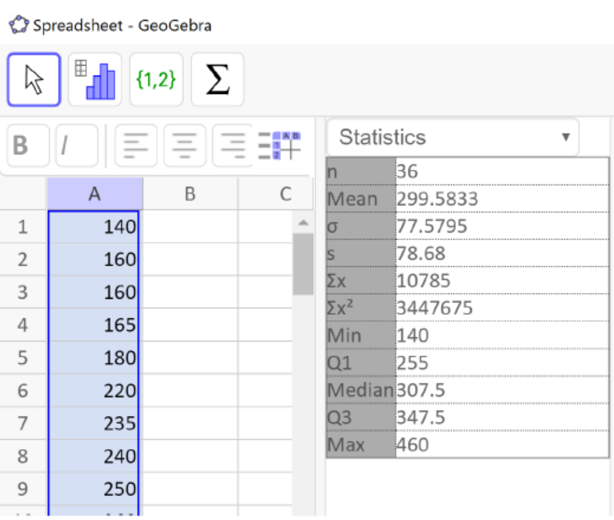
$140 $160 $160 $165 $180 $220 $235 $240 $250

$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

Solution. As we found before, here is the GeoGebra output. The last five entries of the summary statistics are the five-number summary. Remember to label all of your statistics with units.

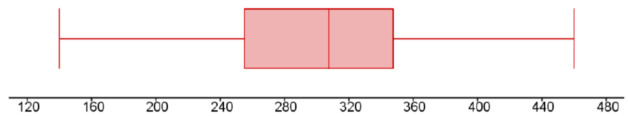


The five-number summary is

Min	Q_1	Median	Q_3	Max
$140	$255	$307.50	$347.50	$460

To draw the boxplot, we will first draw a number line that extends a little beyond the minimum and maximum values, and choose a scale. We decided to draw our number line from $120 to $480, in increments of $40. Then we add a meaningful title and units.

Next, make vertical lines at the first quartile, median and third quartile and connect them to form a box. This is the middle 50% of the data and you might notice that the width of the box is the IQR. Then, extend the "whiskers" out to the minimum and maximum values. Note that a boxplot does not have a vertical scale and the height of the box does not matter. Our boxplot looks like this:

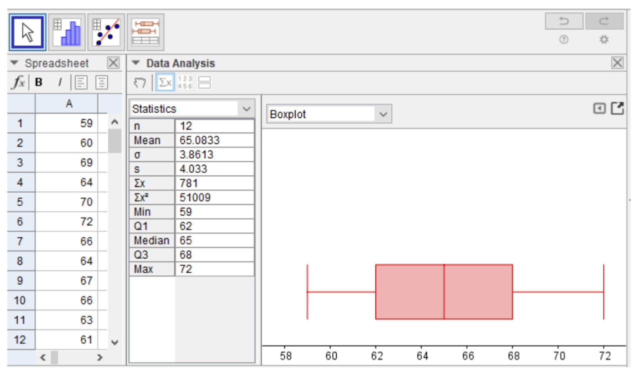


GeoGebra will also draw boxplots for us. We enter and select the data values like we have done before and select One Variable Statistics. This brings up the graphics window with a histogram by default. Use the drop-down menu to select the boxplot. We also click on $\sum x$ to show the summary statistics. □

Example 3.4.7 We will continue with our height data from the 12 people who identify as women. Find the five-number summary and create a boxplot using GeoGebra.

Solution. 59 60 69 64 70 72 66 64 67 66 63 61

Following the steps above we have the following GeoGebra output. The last five entries in the statistics table are the five-number summary and we have the boxplot on the right.

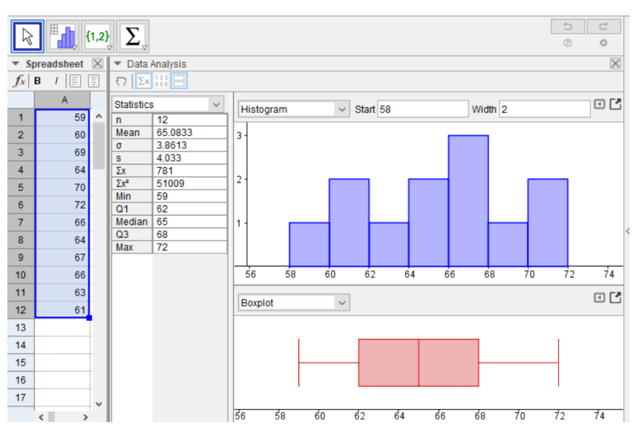


Here is the five-number summary:

Min	Q_1	Median	Q_3	Max
59in	62in	65in	68in	72in

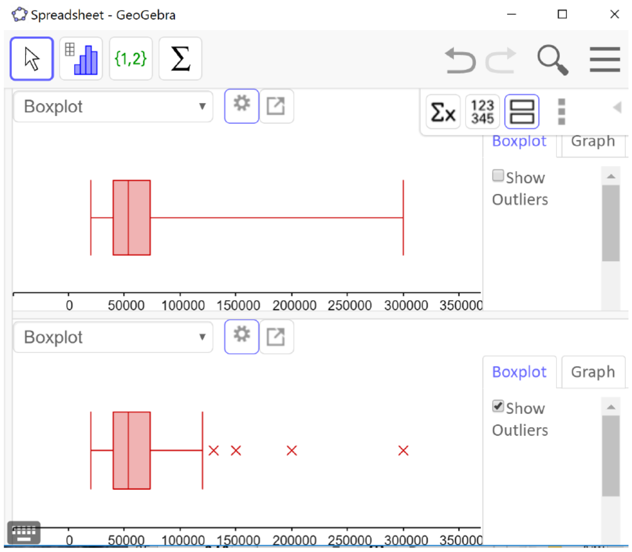
For this data, the two sides of the box and the two whiskers are approximately the same width. This suggests that the distribution is symmetric. We can verify this by noticing that the mean is approximately equal to the median.

The boxplot can tell us the shape of the distribution, but we cannot tell how many peaks the data has. For that we need a histogram. We can see the histogram and boxplot together by selecting the icon that looks like two rectangles stacked, or an = sign.



Now we have a full picture of this data. □

The default boxplot in GeoGebra is called a ***modified box plot***, which shows the data values that are outliers with an X but requires a few more steps to make by hand. To change from the modified box plot to a regular box plot, click on the left pointing arrow in the boxplot window (downloaded version) or the settings wheel (online version) for options, and uncheck "show outliers." The output window below shows two side-by-side boxplots (the regular boxplot on top and the modified boxplot on the bottom) illustrating the distribution of the annual salaries for 50 randomly selected full-time workers in the Portland Metro area.



From the upper boxplot we can see that this distribution is skewed to the right and the upper quarter of the data is very spread out. It is natural to think of the data values as being evenly spread out in each region, but that is quite often not the case. From the lower boxplot we can see that there are 4 data values that are considered outliers and how far away the last data value is from the others. This is why it is useful to show outliers on a boxplot.

3.4.6 Modified Boxplot (Optional)

If you are curious to know how a modified boxplot is made, we will explain it briefly. There is a rule called the ***1.5∗IQR*** rule to determine which points are considered outliers. An outlier is a point that is more than 1.5 IQRs away from the middle 50% of the data (the box in the boxplot). We know how to calculate the IQR and then we multiply that by 1.5. We subtract that from Q1 to find the ***lower fence*** and add that value to Q3 to find the ***upper fence***. Values beyond the fences are considered outliers and are drawn with an X or a star. Then we draw the whiskers of a modified box plot to the furthest data value inside the fence on each side.

3.4.7 Percentiles

Back when we were finding the median, we mentioned that the median is also called the 50th percentile, because 50% of the data values lie below it. We can define any ***percentile*** as the data value with that percentage of values below it. Since we have found the quartiles, we can also identify the 25th and 75th percentiles for our data. Q1 is the 25th percentile because 25% of the data values lie below it and Q3 is the 75th percentile because 75% of the data values lie below it.

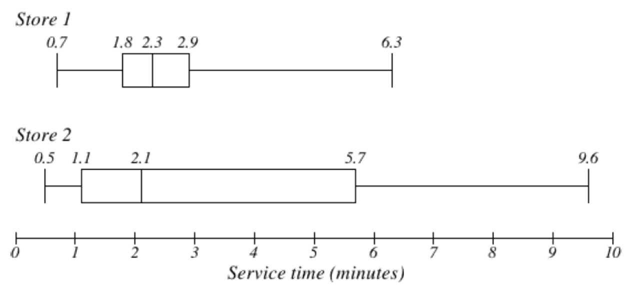
Percentiles are used when comparing the growth of children to the population and in the results of standardized tests, such as the SAT test. If a person scored in the 83rd percentile, that means they scored higher than 83% of the people who took the test.

3.4.8 Comparing Distributions

Box plots and percentiles are particularly useful for comparing data from two populations.

Example 3.4.8 The box plots of service times for two fast-food restaurants are shown below. Compare the length of time to get served at the two restaurants. Which one should you go to if you are in a hurry?

Solution.



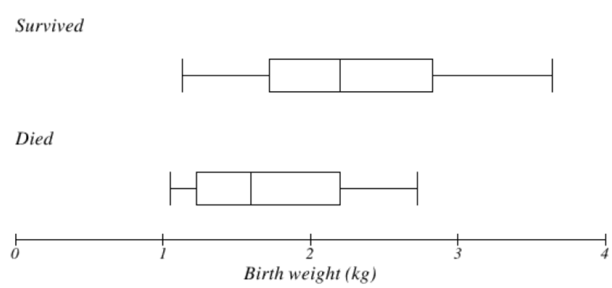
Store 2 has a slightly shorter median service time (2.1 minutes vs. 2.3 minutes), but the service times are less consistent, with a wider spread of the data.

The 75th percentiles are 2.9 and 5.7 minutes. That means 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? That depends upon your opinion about luck – 25% of customers at store 2 had to wait between 5.7 and 9.6 minutes. □

Example 3.4.9 The boxplots below show the 5-number summaries of the birth weights, in kilograms, of infants with severe idiopathic respiratory distress syndrome (SIRDS)[1]. The boxplots are separated to show the birth weights of infants who survived and those that did not. What can we conclude from this data?

Solution.



Comparing the two groups, the boxplot 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 weight of infants that survived is about the same as the third quartile of the infants that died. Similarly, we can see that the 25th percentile of the survivors is larger than the 50th percentile 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 plots give us a quick, though informal, way to determine that birth weight is quite likely linked to the survival of infants with SIRDS. □

[1] 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

3.4.9 Z-Scores

Have you ever heard the saying that you can't compare apples and oranges? It turns out that you can - provided we standardize their measures first!

We will be using the standard score called a Z-score, which is a method commonly used with unimodal and symmetric distributions (called ***normal*** or ***nearly normal distributions***). Z-scores may be used with any data, but if the distribution is skewed, then the distribution of Z-scores will also be skewed.

To calculate the Z-score for a data value, we find out how far away from the mean it is by subtracting. Then we divide by the standard deviation to see how many standard deviations that is. Thus, the ***Z-score*** of a data value is the number of standard deviations it is away from the mean.

Z-score.

$$Z = \frac{\text{data value - mean}}{\text{standard deviation}}$$

Be sure to calculate the difference first, then divide

If a data value is above the mean, its Z-score will be positive. If a data value is below the mean, its Z-score will be negative. Therefore, if a data value is one standard deviation above the mean, its Z-score is +1. If it is 2.5 standard deviations below the mean, its Z-score is -2.5. Note that the units of Z-scores are standard deviations, not the units of the data values.

We can use Z-scores to determine the relative unusualness of a data value with respect to its own distribution. That is what allows us to compare two unlike items. The convention in statistics is to say that a data value is ***unusual*** if it is more than 2 standard deviations from the mean, or in other words, if its Z-score is less than -2 or greater than +2.

Example 3.4.10 The oranges at a local grocery store have a mean diameter of 5.8 inches and a standard deviation of 1.2 inches. The apples, on the other hand, have a mean diameter of 4.2 inches and a standard deviation of 0.6 inches.

Ali closes their eyes and selects and apple and an orange. When they look at both pieces of fruit, they seem small. If the orange has a diameter of 4.2 inches and the apple has a diameter of 3.5 inches, which is smaller relative to their respective piles of fruit?

To determine which fruit is relatively smaller, Ali can find each of their Z-scores.

$$\begin{aligned} Z_{\text{Orange}} &= \frac{4.2 - 5.8}{1.2} = -1.33 \text{ standard deviations} \\ Z_{\text{Apple}} &= \frac{3.5 - 4.2}{0.6} = -1.17 \text{ standard deviations} \end{aligned}$$

By convention, Z-scores are rounded to two decimal places, so we see that the orange is 1.33 standard deviations below its mean and the apple is 1.17 standard deviation below its mean. The orange is therefore smaller relative to its distribution since its Z-score is less than the apple's Z-score.

We can also see from the Z-scores that neither fruit has an unusually small diameter since each piece of fruit is less than 2 standard deviations from its mean.

We can also find Z-scores using a spreadsheet with this formula:

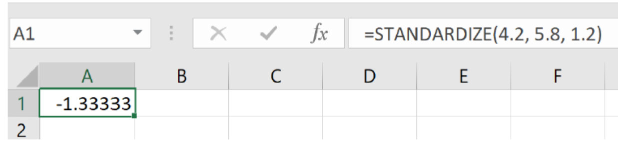
```
=STANDARDIZE(data value, mean, standard deviation)
```

To verify our apple and orange Z-scores, we would write:

Apple:

```
=STANDARDIZE(4.2, 5.8, 1.2)
```

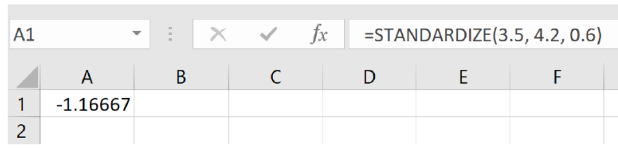
and we get -1.33 standard deviations



Orange:

```
=STANDARDIZE (3.5, 4.2, 0.6)
```

and we get -1.17 standard deviations



□

Example 3.4.11 The mean weight of men over the age of 20 is 195.7 pounds[2] with a standard deviation of 29.8 pounds. The mean weight of domestic cats is 8.6 pounds with a standard deviation of 1.2 pounds. (The standard deviation for men's weights is estimated. The cat's mean weight is based on ideal cat weight and the standard deviation is approximate).

At his peak, Andre the Giant, the 7-foot-4-inch French professional wrestler and actor, weighed 520 pounds. When Georgie the cat was at his peak he weighed 24 pounds. Who was more giant – Andre the Giant or Georgie the cat?

Since the weights of cats and men cannot be compared directly, we will need to calculate the Z-scores.

$$Z_{\text{Andre}} = \frac{520 - 195.7}{29.8} = 10.88 \text{ standard deviations.}$$
$$Z_{\text{Georgie}} = \frac{24 - 8.6}{1.2} = 12.83 \text{ standard deviations.}$$

Using the standardize function, we would write:

Andre:

```
=STANDARDIZE (520, 197.5, 29.8)
```

which gives us 10.88 standard deviations.

Georgie:

```
=STANDARDIZE (24, 8.6, 1.2)
```

which gives us 12.83 standard deviations.

Since both Z-scores are greater than 2 standard deviations, both weights are extremely unusual. However, since the Z-score for Georgie's weight is larger, he is even more giant than Andre the Giant. □

3.4.10 Exercises

1. A group of diners were asked how much they would pay for a meal. Their responses were: \$7.50, \$25.00, \$10.00, \$10.00, \$7.50, \$8.25, \$9.00, \$5.00, \$15.00, \$8.00, \$7.25, \$7.50, \$8.00, \$7.00. \$12.00.

a. Using your mean from Exercise 3.3.7.1, find the standard deviation of this data. Explain what the mean and standard deviation tell you about how much the group of diners would pay for a meal.

b. Calculate the five-number summary for this data.

c. Calculate the range and IQR for this data.

d. Create a boxplot for the data.

2. The amount of commercials in an hour of television varies by channel. The total length (in minutes) of all commercials from 8 pm to 9 pm in for some selected broadcast and cable channels on a weekday

[2] 2016 CDC Report. https://www.cdc.gov/nchs/data/series/sr_03/sr03_039.pdf. The study included all ethnicities but the report does not say whether transgendered men were included.

evening were: 10, 12.75, 7, 9, 9.75, 6.5, 12.5, 12.5, 8.75, 17, 10.5, 2

a. Using your mean from Exercise 3.3.7.2, find the standard deviation of this data. Explain what the mean and standard deviation tell you about how much the group of diners would pay for a meal.

b. Calculate the five-number summary for this data.

c. Calculate the range and IQR for this data.

d. Create a boxplot for the data.

3. 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.

a. Using your mean from Exercise 3.3.7.3, find the standard deviation of this data. Explain what the mean and standard deviation tell you about how much the group of diners would pay for a meal.

b. Calculate the five-number summary for this data.

c. Calculate the range and IQR for this data.

d. Create a boxplot for the data.

4. You weigh 9 Oreo cookies, and you find the weights (in grams) are: 3.49, 3.51, 3.51, 3.51, 3.52, 3.54, 3.55, 3.58, 3.61

a. Using your mean from Exercise 3.3.7.4, find the standard deviation of this data. Explain what the mean and standard deviation tell you about the weights of these Oreo cookies.

b. Calculate the five-number summary for this data.

c. Calculate the range and IQR for this data.

d. Create a boxplot for the data.

5. The following table shows the cost of purchasing a car at a local dealership. Some of the cars sold were new and some were used.

a. Find the standard deviation of this data. Explain what the mean and standard deviation tell you about how much the cars are selling for.

b. Calculate the five-number summary for this data.

c. Calculate the range and IQR.

d. Create a boxplot for the data.

Cost (Thousands of dollars)	Frequency
15	3
20	7
25	10
30	15
35	13
40	11
45	9
50	7

6. As part of a study of email, a researcher counted the length of 34 emails. The lengths of the emails are shown below, rounded to the nearest thousand characters (so a length 0 means that the numbers of characters rounded to 0, not that the message was blank).

a. Find the standard deviation of this data. Explain what the mean and standard deviation tell you about the length of the emails.

b. Calculate the five-number summary for this data.

c. Calculate the range and IQR.

d. Create a boxplot for the data.

Length of an email (Thousands of characters)	Frequency
0	4
1	5
2	2
3	3
4	3
5	1
6	3
7	3
8	0
9	3
10	3
11	2
12	0
13	0
14	2

7. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment. Suppose that a new cancer drug is currently under study. 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 cancer patients throughout their treatment. The following data (in months) are collected.

a. Find the standard deviation of each group.

b. Calculate the 5-number summary for each group.

c. Calculate the range and IQR for each group.

d. Create side-by-side boxplots and compare and contrast the two groups.

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, 34

Researcher 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

8. The US Census Bureau, in addition to counting the population of the US every 10 years, conducts yearly informational surveys, such as the American Community Survey (ACS). For the 2012 ACS, a randomly chosen group of 20 respondents (10 males, 10 females) answered a question about their incomes.

Males: \$53,000; \$70,000; \$12,800; 30,000; \$4,500; \$42,000; \$48,000; \$60,000; \$108,000; \$11,000

Females: \$1,600; \$1,200; \$20,000; \$25,000; \$670; \$29,000; \$44,000; \$30,000; \$5,800; \$50,000

a. Find the standard deviation of each group.

b. Calculate the 5-number summary for each group.

c. Calculate the range and IQR for each group.

d. Create side-by-side boxplots, and compare and contrast the two groups.

9. An experiment compared the ability of three groups of participants to remember briefly-presented chess positions. The data are shown below. The numbers represent the average number of pieces correctly remembered from three chess positions.

a. Find the standard deviation of each group.

b. Calculate the 5-number summary for each group.

c. Calculate the range and IQR for each group.

d. Create side-by-side boxplots and compare and contrast the two groups.

Non-players	Beginners	Tournament Players
22.1	32.5	40.1
22.3	37.1	45.6
26.2	39.1	51.2
29.6	40.5	56.4
31.7	45.5	58.1
33.5	51.3	71.1
38.9	52.6	74.9
39.7	55.7	75.9
39.7	55.7	75.9
43.2	55.9	80.3
43.2	57.7	85.3

10. There is evidence that smiling can attenuate judgments of possible wrongdoing. This phenomenon termed the "smile-leniency effect" was the focus of a study by Marianne LaFrance & Marvin Hecht in 1995[3]. The following data are measurements of how lenient the sentences were for three different types of smiles and one neutral control. The same subject was used for all of the conditions so that may affect the results. The second column is a continuation of the first column.

a. Find the standard deviation for each type of smile and the neutral control.

b. Calculate the 5-number summary for type of smile and the neutral control.

c. Calculate the range and IQR for each type of smile and the neutral control.

d. Create side-by-side boxplots and compare and contrast the four groups.

False Smile	Felt Smile	Miserable Smile	Nuetral Control
2.5	7	5.5	2
5.5	3	4	4
6.5	6	4	4
3.5	4.5	5	3
3	3.5	6	6
3.5	4	3.5	4.5
6	3	3.5	2
5	3	3.5	6
4	3.5	4	3
4.5	4.5	5.5	3
5	7	5.5	4.5
5.5	5	4.5	8
3.5	5	2.5	4
6	7.5	5.5	5
6.5	2.5	4.5	3.5
3	5	3	4.5
8	5.5	3.5	6.5
6.5	5.5	8	3.5
8	5	5	4.5
6	4	7.5	4.5
6	5	8	2.5
3	6.5	4	2.5
7	6.5	5.5	4.5
8	7	6.5	2.5
4	3.5	5	6
3	5	4	6
2.5	3.5	3	2
8	9	5	4
4.5	2.5	4	5.5
5.5	8.5	4	4
7.5	3.5	6	2.5
6	4.5	8	2.5
9	3.5	4.5	3
6.5	4.5	5.5	6.5

11. Make up two data sets with 5 numbers each that have:

a. The same mean but different standard deviations.

b. The same standard deviation but different means.

12. Make up two data sets with 7 numbers that each have:

a. The same IQR but different medians.

b. Different IQRs but the same medians.

13. The side-by-side boxplots show salaries for actuaries and CPAs.

a. Estimate the 25th, 50th and 75th percentiles for CPA and actuary salaries.

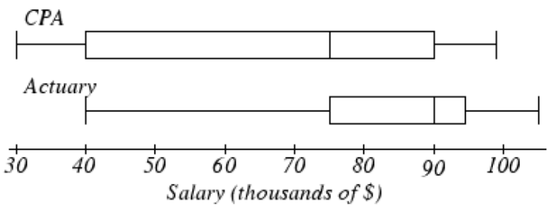
b. Deshawn makes the median salary for an actuary. Kelsey makes the first quartile salary for a CPA. Who makes more money? How much more?

---

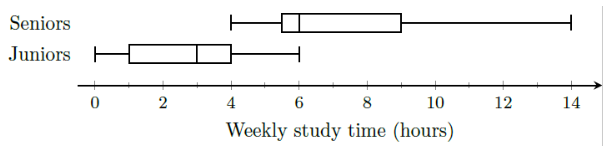
[3] LaFrance, M., & Hecht, M. A. (1995) Why smiles generate leniency. Personality and Social Psychology Bulletin, 21, 207-214. Adapted from www.onlinestatbook.com, by David M. Lane, et al, used under CC-BY-SA 3.0.

c. What percentage of actuaries make more than the median salary of a CPA?

d. What percentage of CPAs earn less than all actuaries?



14. Fifty juniors and fifty seniors at a local high school were surveyed to find out how many hours per week they spend studying. The side-by-side boxplots the weekly study times for those high school juniors and seniors.



a. Estimate the 25th, 50th, and 75th percentiles for weekly study time for high school juniors and seniors.

b. Olivia studies the maximum number of weekly study hours for a junior. Lucy studies the first quartile weekly study time for a senior. Who studies more, and by how many hours?

c. What percentage of juniors study between the minimum and median weekly study times for seniors?

d. What percentage of seniors study more than the third quartile weekly study time for juniors?

15. Suppose you buy a new car whose advertised gas mileage is 25 mpg (miles per gallon). After driving the car for several months, you find that you are getting only 21.4 mpg. You phone the manufacturer and learn that the standard deviation of gas mileage for cars of that model is 1.15 mpg.

a. Find the Z-score for the gas mileage of your car.

b. Does it appear that your car is getting unusually low gas mileage? Explain your answer using your Z-score.

16. According to a local marathon club, the mean finishing time for a marathon is 274 minutes, with a standard deviation of 63 minutes.

a. If I can run a marathon with a finishing time of 170 minutes, find the Z-score for my marathon time.

b. Is my marathon finishing time of 170 minutes unusually fast? Explain your answer using the Z-score.

17. This data is a sample of the average number of hours per year that a driver is delayed by road congestion in 11 cities: 56, 53, 53, 50, 46, 45, 44, 43, 42,40, 36

a. Find the mean and the standard deviation, including units.

b. What is the Z-score for the city with an average delay time of 42 hours per year?

18. In a survey of 12 companies recruiting for recent college graduates, they reported the following numbers of job applicants per job posting: 123, 123, 134, 127, 115, 122, 125, 101, 130, 143, 110, and 122.

a. Find the mean and standard deviation, including units.

b. What is the Z score for the company with 143 job applicants per job posting?

19. You scored an 89 on a math test where the class mean and standard deviation are 75 points and 7 points respectively. You scored a 65 on an English test where the mean and standard deviation are 53 points and 4 points, respectively. In which class did you do better? Explain your answer using Z-scores.

20. The mean running time for comedy movies is 139 minutes, with a standard deviation of 39.7 minutes. For action movies, the mean running time is 159 minutes, with a standard deviation of 26.2 minutes. A recent comedy movie had a running time of 102 minutes, while an action movie playing at the same theatre had a running time of 129 minutes. Which movie is shorter compared to other movies in the same genre? Explain your answer using Z-score.

21. Poe, the Clydesdale horse has a world record breaking height of 20.2 hands. All Clydesdale horses have a mean height of 16.5 hands and a standard deviation of 1.85 hands. The last Great Dane to hold the world record for dog height was Gibson who was 107 cm tall. Great Danes have a mean height of 81 cm and a standard deviation of 13 cm. Which animal is taller compared to their respective breed? Explain your answer using Z-scores.

3.5 Chapter 3 Review

Review Exercises

1. Portland Community College serves nearly 73,000 full-time and part-time students in the greater Portland area at four main campuses (SE, Cascade, Sylvania, and RC). Student Affairs would like to know how students get to campus. They randomly select 250 students from each of the four main campus and ask them how they got to classes on campus. The following are the results of their survey:

- Public Transportation: 435
- Driving: 475
- Biking: 65
- Walking: 30

a. Identify the population and state its size.

b. Identify the sample and state its size.

c. What sampling method was used?

d. What type of data was collected?

e. Give the statistic for the percentage of students who use public transportation.

2. CNN conducted a survey of 500 American adults. 62% of those surveyed answered yes to the question, "Do you favor a law to ban the sale of assault weapons and semiautomatic rifles?" The reported margin of error was $\pm 4\%$.

a. What population is being studied?

b. What is the sample?

c. What type of data is this?

d. Is the 62% reported in the problem an example of a statistic or a parameter?

e. What is the confidence interval? Is the confidence interval about the statistic or the parameter?

f. Explain what the confidence interval tells you.

3. A survey of 265 PCC students found that 23%, plus or minus 4% prefer to study at the library.

a. What population is being studied?

b. What type of data was collected?

c. Is the reported 23% a statistic or a parameter?

d. What is the margin of error?

e. What is the confidence interval?

f. Explain what the confidence interval tells you.

4. Identify the sampling method. Just the name will suffice.

a. Researchers select every 5th customer who walks into the store to take a survey.

b. Raffle tickets are distributed and collected in a bag, where they are mixed and ten are drawn for prizes.

c. I asked the shoppers near me in the shoe department what size they wear.

d. An IRS auditor randomly selects 25 taxpayers in each filing status (single, head of household, married filing jointly, and married filing separately).

5. Identify the most relevant source of bias in each situation.

a. An opinion poll is posted on Facebook and Twitter asking how you are most likely to vote for in the next election.

b. Keller Auditorium ask all the people in the front three rows if they enjoyed the Broadway play.

c. To determine opinions on voter support for a downtown farmers market, a surveyor randomly questions people working close to the park where the farmers market would be.

d. A survey asks people to report the number of hours they work out each week.

e. A survey randomly calls people on their landlines and ask them if they would support a school bond measure in the next election.

6. Identify whether each situation describes an observational study or an experiment. If it is an experiment.

a. Subjects are asked to run a mile and record their time.

b. Fifty students were asked to go to a quiet space in the library to memorize a poem. Fifty students were asked to go to a noisy location in the cafeteria to memorize the poem. Each student recorded how much time it took to memorize the poem.

7. For the clinical trial of a migraine drug, subjects were randomly divided into two groups. The first received an inert pill, while the second received the test medicine. Patients were not aware of which group they were in. After one month, patients reported how many migraines they experienced.

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?

8. In a recent study[1], 380 high risk adolescents involved in the juvenile justice system were recruited to test an app designed to increase mindfulness and reduce substance use. Participants were randomly and equally assigned to use the app (Rewire) or receive services as usual from the Department of Youth Services. Participants were assessed to determine a baseline for substance use at the beginning of the study, and were asked to complete follow up assessments after 1 and 3 months. Assessments consisted of online surveys asking about substance use, emotion regulation, family demographics, and mindfulness practices. Urine samples were collected at each interview to verify self-reported substance use.

a. Describe the treatment group.

b. Describe the control group (if there is one).

c. Is this study blind, double-blind, or neither? Explain

d. Is this best described as an experiment, a controlled experiment, or a placebo-controlled experiment?

9. In a 2010 survey, US teens aged 12-18 were asked what their favorite movie genre was. The results are shown below.

- Action: 351
- Adventure: 171
- Comedy: 651
- Drama: 389

[1] https://clinicaltrials.gov/ct2/show/NCT04027075?map_cntry=US&map_state=US%3AOR&rank=1

- Horror: 287
- Romance: 107
- Undecided: 51

a. What is the implied population?

b. How many people were sampled?

c. What type of data is this?

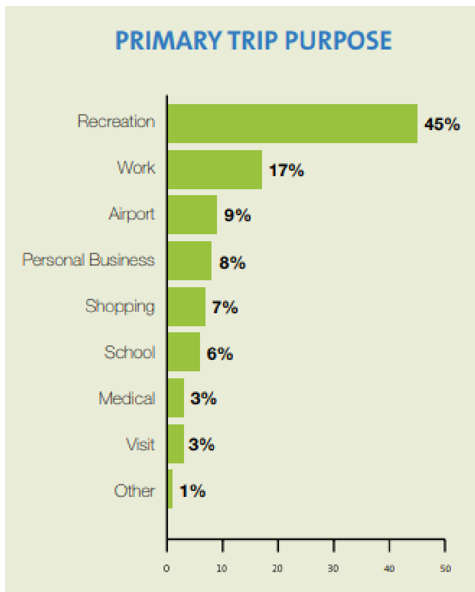
d. Create a relative frequency bar chart of the results.

e. Create a pie chart of the results.

f. Explain the advantages/disadvantages of the two charts.

g. What is the statistic for the percentage of teens whose favorite movie genre is horror?

10. A survey of 5325 Portland residents was conducted to determine the primary purpose of using TriMet. The results are shown below.



a. How many people use TriMet for personal business?

b. How many people use TriMet to get to the airport?

11. A group of college students were asked what the price of gas would need to be before they would start using public transportation to get to school instead of driving. Their responses in \$/gallon are listed below:

5.25, 5.00, 4.25, 3.75, 5.00, 4.50, 3.95, 3.75, 5.75, 4.75, 3.25, 3.75, 4.75, 5.00, 8.95

a. Find the mean and median. Round to two decimal places and include units.

b. Based on the mean and median, would you expect the distribution to be symmetric, skewed left, or skewed right? Explain.

c. Find the standard deviation. Round to two decimal places and include units.

d. Calculate the z-scores for the responses of \$3.25 and \$8.95. Are either of these values unusual?

e. Determine the 5-number summary for the data.

f. What is the range and IQR of the data set? Round to two decimal places and include units.

g. Use the 5-number summary to construct a box plot.

12. The following is a sample of scores from a recent Math 105 exam:

32, 71, 72, 73, 73, 73, 76, 77, 78, 78, 79, 86, 88, 88, 88, 94, 94, 99

a. Find the mean of the data. Round to one decimal place if necessary.

b. Find the median of the data. Round to one decimal place if necessary.

c. Just comparing the mean and the median, do you expect the distribution to be skewed left, skewed right, or symmetric. Explain.

d. Find the standard deviation of the data. Round to one decimal place if needed.

e. Explain what the mean and standard deviation tell you about the sampled test scores.

f. Is the score of 99 unusual? Use z-scores to support your claim.

g. Find the 5-number summary.

h. Use the 5-number summary to create a box plot.

i. Create a histogram of the data. Start your scale at 0, and use a bin size of 10.

j. Describe the shape of the distribution. Be sure to address all three characteristics (modality, symmetry, and outliers).

13. The following table shows the cost of purchasing a car at a local dealership. Some of the cars sold were new and some were used.

Cost (Thousands of dollars)	Frequency
12	6
15	7
18	12
22	10
30	12
32	11
40	6
45	6

a. Find the mean and standard deviation of the data. Round to two decimal places and include units.

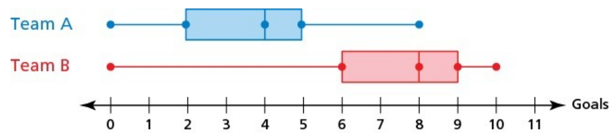
b. Explain what standard deviation tell you about how much cars are selling for at this dealership.

c. Determine the five-number summary.

d. What is the range and IQR?

e. Use the five-number summary to construct a boxplot of the data.

14. The double box-and-whisker plot[2] shows the goals scored per game by two soccer teams during a 25 game season.



a. Estimate the 25th, 50th and 75th percentiles for Team A and Team B goals.

b. What is the median number of goals for Team A? Team B?

c. What percentage of the goals for Team B is more than the maximum number of Team A?

d. What Team data is more symmetric?

e. What is the shape of the distribution for Team B?

15. Suppose you buy a new car whose advertised gas mileage is 35 mpg (miles per gallon). After driving the car for several months, you find that you are getting only 30.4 mpg. You phone the manufacturer and learn that the standard deviation for that model is 1.35 mpg.

a. Find the z-score for the gas mileage of your car.

b. Does it appear that your car is getting unusually low gas mileage? Explain your answer using your z-score.

16. This data is a sample of the average number of minutes per week that a driver is delayed by road congestion in 13 cities:

66, 55, 53, 50, 36, 45, 34, 43, 52, 40, 76, 45, 63

a. Find the mean and the standard deviation, including units.

b. What is the z-score for the city with an average delay time of 42 hours per week?

c. Is an average delay time of 42 hours per week unusual? Explain using the calculated z-score.

[2]Source: https://brainly.com/question/11978942

3.6 The Normal Distribution

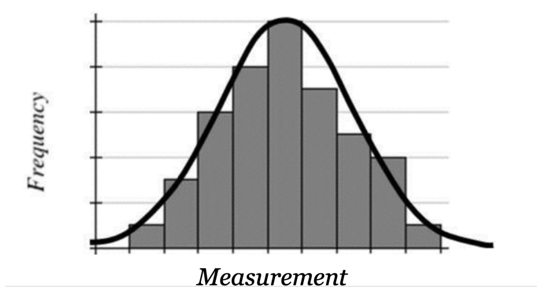
Objectives: Section 3.6 The Normal Distribution

Students will be able to:

- Explain the properties of the Normal distribution
- Calculate probabilities using the Empirical Rule
- Calculate and interpret Z-scores
- Calculate and interpret the bounds of a 95% confidence interval given the point estimate and the standard error

3.6.1 Introduction to The Normal Distribution

There is a pattern or shape that comes up frequently when graphing data called the ***Normal distribution***. You might be familiar with it as the ***bell curve***. When we take measurements like the lengths of newborn babies or the weights of potato chip bags on an assembly line there are usually more measurements near the mean and fewer as we get further away from the mean on each side. We often see histograms with this shape:

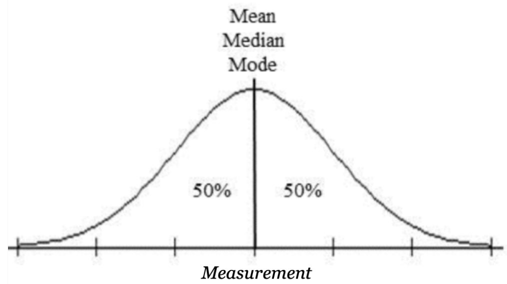


Notice how the shape is unimodal and symmetric with larger and smaller values getting less and less common as they get further from the mean. When we take samples from a population we might not get an exact Normal distribution but it is often close. We call this approximately Normal or nearly Normally distributed. The curve drawn over the histogram shows that the data are nearly Normal. Earlier in this chapter we learned about the ***mean***, ***median*** and ***mode*** as measures of center and the ***standard deviation*** as a measure of variation or spread. The mean is often written with the greek letter mu, μ. The standard deviation is usualy represented by the lowercase Greek letter sigma, σ.

3.6.2 Properties of the Normal Distribution

Regardless of the data values, the Normal distribution has these properties:

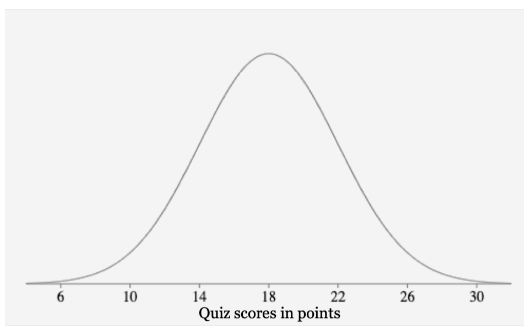
- The shape is unimodal and symmetric around the center.
- The mean, median and mode are approximately the same value in the center.
- Due to the symmetry, approximately 50% of the data values are below the mean, and approximately 50% of the data values are above the mean.



- Almost all of the area under a Normal curve is within 3 standard deviations of the mean, so we use the standard deviation as our scale.
- Data within two standard deviations of the mean are considered typical or usual. Data beyond two standard deviation are considered unusual. Data beyond 3 standard deviations are considered rare.

Now let's look at an example where we draw and label a Normal distribution given the mean and standard deviation[1].

Example 3.6.1 The test scores on a math quiz are approximately normally distributed with a mean of 18 points and a standard deviation of 4 points. Draw the Normal distribution and label the axis using the standard deviation.



We start by drawing a Normal curve and the horizontal axis. Then we place the mean of 18 points in the center of the graph and make 3 marks on each side, ending where the curve gets close to the axis.

Each mark represents one standard deviation. As we move to the right of the mean we will add the standard deviation of 4 points to get 22, 26 and 30 points. These are the values that are 1, 2 and 3 standard deviations above the mean.

$$\text{Mean} + 1 \text{ standard deviation} = 18 + 4 = 22$$
$$\text{Mean} + 2 \text{ standard deviations} = 18 + 2(4) = 26$$
$$\text{Mean} + 3 \text{ standard deviations} = 18 + 3(4) = 30$$

Then we move to the left of the mean we subtract the standard deviation of 4 from 18 to get 14, 10 and 6 points. These are the values that are 1, 2 and 3 standard deviations below the mean.

$$\text{Mean} - 1 \text{ standard deviation} = 18 - 4 = 14$$
$$\text{Mean} - 2 \text{ standard deviations} = 18 - 2(4) = 10$$
$$\text{Mean} - 3 \text{ standard deviations} = 18 - 3(4) = 6$$

[1]All of the Normal drawings in this section were created with the online probability calculator at onlinestatbook.com/2/calculators/normal_dist.html which is in the public domain.

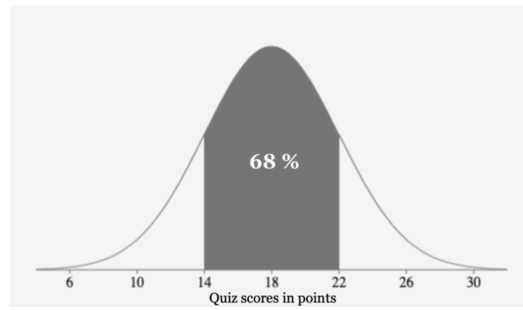
□

3.6.3 The Empirical Rule

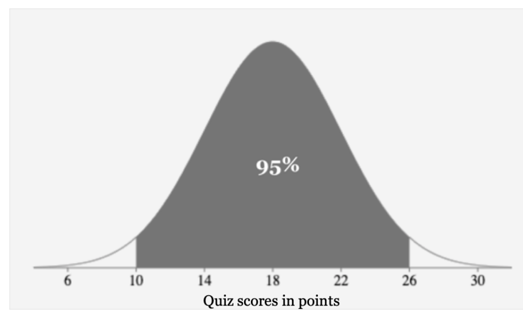
No matter what the mean and standard deviation are for a Normal Distribution, they always have the same shape or distribution of the data. We can describe this with the ***Empirical Rule*** which is also called the ***68-95-99.7 Rule***.

The numbers in the 68-95-99.7 rule describe the percentage of data or area within 1, 2 and 3 standard deviations of the mean. Let's look at our previous example with scores on a math quiz that are approximately normally distributed with a mean of 18 points and a standard deviation of 4 points.

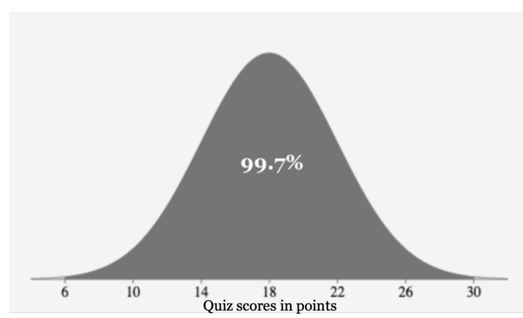
According to the Empirical rule, about 68% of all the data values fall within one standard deviation of the mean, or between 14 and 22 points.



About 95% of all the data values fall within two standard deviation, or between 10 and 26 points.



About 99.7% of all the data values fall within three standard deviations, or between 6 and 30 points.



Now you can try this next example:

Example 3.6.2 The weights of black lab dogs are approximately normally distributed with a mean of 78 pounds and a standard deviation of 7 pounds.

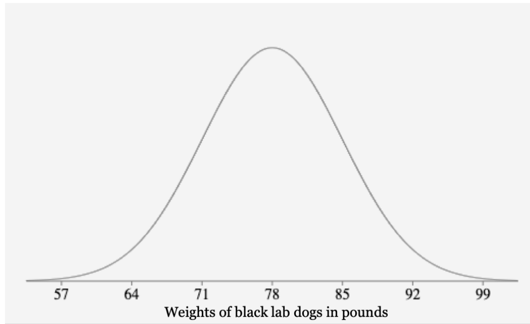
a. Draw and label a sketch of this distribution.

b. Use the Empirical Rule to complete following statements.

- Approximately 68% of black labs weigh between _________ and _________ pounds.
- Approximately 95% of black lab dogs weigh between _________ and _________ pounds.
- Approximately 99.7% of black lab dogs weigh between _________ and _________ pounds.

Solution.

a. We start by drawing the Normal curve and the horizontal axis, labeling 3 standard deviations on each side.

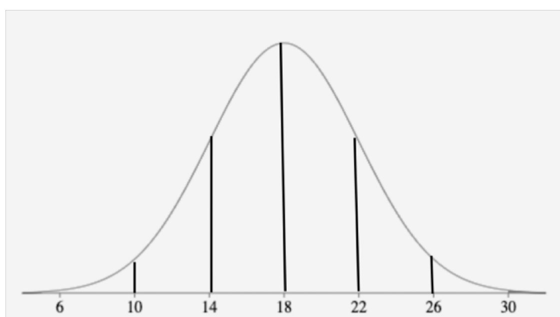


b.
- Approximately 68% of black labs weigh between 71 and 85 pounds.
- Approximately 95% of black labs weigh between 64 and 92 pounds.
- Approximately 99.7% of black labs weigh between 57 and 99 points.

□

3.6.4 Using the Empirical Rule to Calculate Approximate Probabilities

Now that we have learned about the Empirical rule we can use it to find approximate probabilities. First we will calculate the percentage in each segment of the Normal distribution. Returning to our example of quiz scores with a mean of 18 points and a standard deviation of 4 points, we can divide the curve into segments by drawing a line at each standard deviation.



Quiz scores in points

Now, since we know that approximately 68% of the values are between 14 and 22 points, and the graph is symmetric, we divide 68% by 2 to get:

$$68\% \div 2 = 34\%$$

So each of the middle 2 segments are 34% of the area each.

Next, we know that 95% of the values fall within 2 standard deviations or between 10 and 26 points. That also includes the 68% though, so to get the area of the next two segments, we subtract:

$$95\% - 68\% = 27\%$$

Then since there is one segment on each side, we divide that by 2:

$$27\% \div 2 = 13.5\%$$

So the next two segments from the mean are 13.5% each.

Continuing outward, 99.7% of the values fall within 3 standard deviations of the mean or between 6 and 30 points. Since that includes the 95% we subtract:

$$99.7\% - 95\% = 4.7\%$$

Then we divide that by 2:

$$4.7\% \div 2 = 2.35\%$$

So the next two segments outward are 2.35% each.

Finally, there is one more set, since there is a tiny percentage outside of the 3 standard deviations. The total is 100% so we subtract:

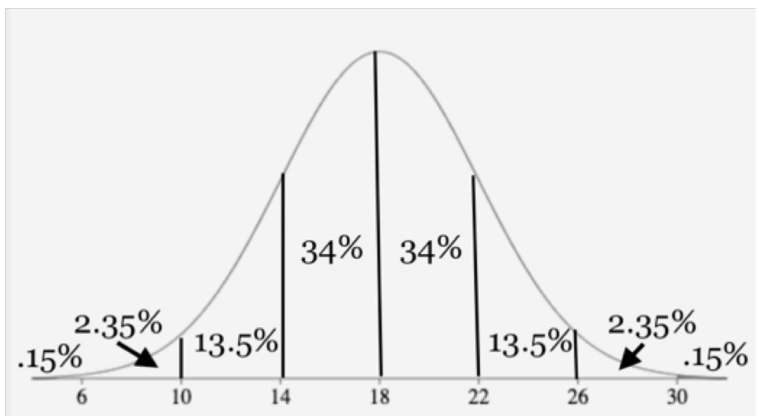
$$100\% - 99.7\% = 0.3\%$$

Then we divide that by 2:

$$0.3\% \div 2 = 0.15\%$$

So the last two outer segments are 0.15% each.

We can summarize all the calculations in this drawing:



Now let's continue our quiz score example to see how to use the Empirical Rule to find probabilities.

Example 3.6.3 For the quiz score data that are approximately normally distributed with a mean of 18 points and a standard deviation 4 points, use the 68-95-99.7% rule to calculate the following probabilities:

a. What percentage of students earned scores between 22 and 30 points?

b. What percentage of students earned scores above 14 points?

c. What percentage of students earned a score below 14 or above 22 points?

Solution

a. Using the diagram above, we add the two segments between 22 and 30 to get

$$13.5\% + 2.35\% = 15.85\%.$$

15.8% percent of students earned scores between 22 and 30 points on the quiz.

b. We add all the segments above 14 to get

$$34\% + 34\% + 13.5\% + 2.35\% + 0.15\% = 84\%.$$

84% of students earned scores above 14 points on the quiz.

c. We add all the segments below 14 and above 22 to get

$$0.15\% + 2.35\% + 13.5\% + 13.5\% + 2.35\% + 0.15\% = 32\%.$$

32% of students earned scores below 14 or above 22 points on the quiz. Notice we could have done this in a faster way by taking

$$100\% - 68\% = 32\%.$$

There can be more than one way to get the answer.

□

Now you can try one with this next example.

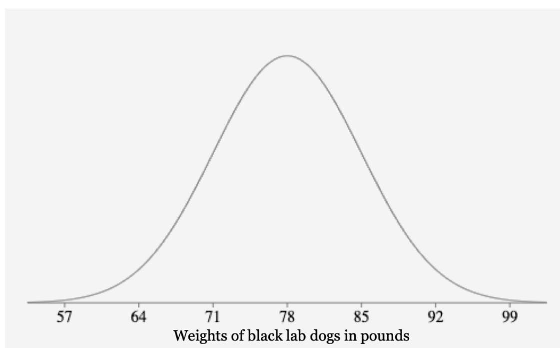
Example 3.6.4 Continuing with the black lab dogs whose weights are approximately normally distributed with a mean of 78 pounds and a standard deviation 7 pounds, use the 68-95-99.7% rule to calculate the following probabilities:

a. What percentage of black labs weigh between 78 and 92 pounds?

b. What percentage of black labs weigh below 71 points?

c. What percentage of black labs weigh below 64 or above 78 points?

Solution. First we need to draw this distribution and label three standard deviations on each side of the mean to determine where these weights fall. That will tell us which segments to add up. Here is our drawing of this distribution from before.



a. Using this Normal curve along with the diagram above we see we need to add the two segments between 78 and 92 pounds to get

$$34\% + 13.5\% = 47.5\%.$$

47.5% percent of black labs weigh between 78 and 92 pounds.

b. We add all the segments below 71 pounds to get

$$13.5\% + 2.35\% + 0.15\% = 16\%.$$

16% of black labs weigh below 71 pounds.

c. We add all the segments below 64 and above 78 pounds to get

$$0.15\% + 2.35\% + 34\% + 13.5\% + 2.35\% + 0.15\% = 52.5\%.$$

52.5% of black labs weigh below 64 or above 78 pounds. Notice we could have done this in a faster way since 78 is the mean we know the percentage above the mean is 50%. We could have added

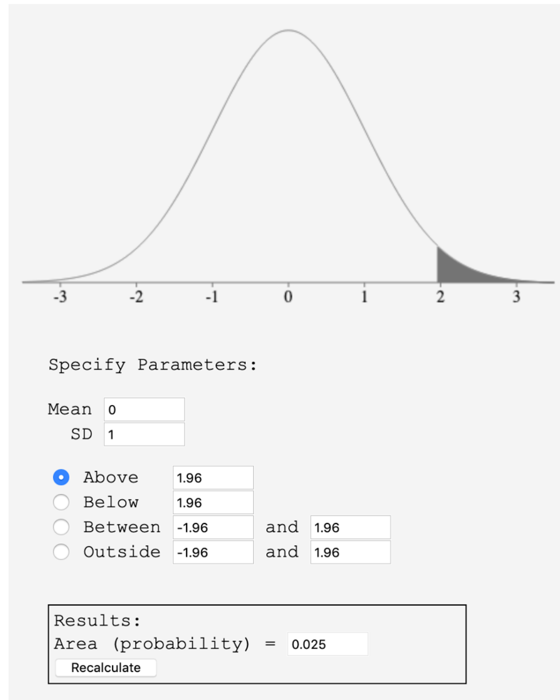
$$0.15\% + 2.35\% + 50\% = 52.5\%.$$

□

3.6.5 Using Technology to Calculate Exact Probabilities

The Empirical rule is used to find approximate probabilities and it only works if the number is one of the labels on our scale. We usually want to calculate an exact answer with any value. We will use technology to calculate probabilities with two different methods: ***Onlinestatbook.com*** and ***spreadsheets***. Check with your class notes or instructor on which technology to use.

To use Onlinestatbook.com, go to the OnlineStatbook Normal calculator and enter the mean and standard deviation. Then select the option that matches the probability you want to find: Above, Below, Between or Outside and enter the probability you want to find. Then click on the Recalculate button.



To use a spreadsheet, we will use the spreadsheet formula

`=NORM.DIST(value, mean, standard deviation, cumulative)`

This function always gives the probability to the left of a number or

$$P(X \leq \text{value}).$$

We can use this formula to find greater than, between and outside probabilities as well. Here are the commands we will use and we will explain them in the next example with visuals.

Probability	Spreadsheet Formula
Below	`=NORM.DIST(value, mean, SD, 1)`
Above	`=1-NORM.DIST(value, mean, SD,1)`
Between	`=NORM.DIST(upper_value, mean, SD, 1)-Norm.Dist(lower_value, mean, SD, 1)`
Outside	`=NORM.DIST(lower_value, mean, SD, 1)+1-NORM.DIST(upper_value,mean, SD, 1)`

Example 3.6.5 For the quiz score data that are approximately normally distributed with a mean of 18 points and a standard deviation 4 points, use technology to calculate the following probabilities:

a. What percentage of students earned a score below 12 points?

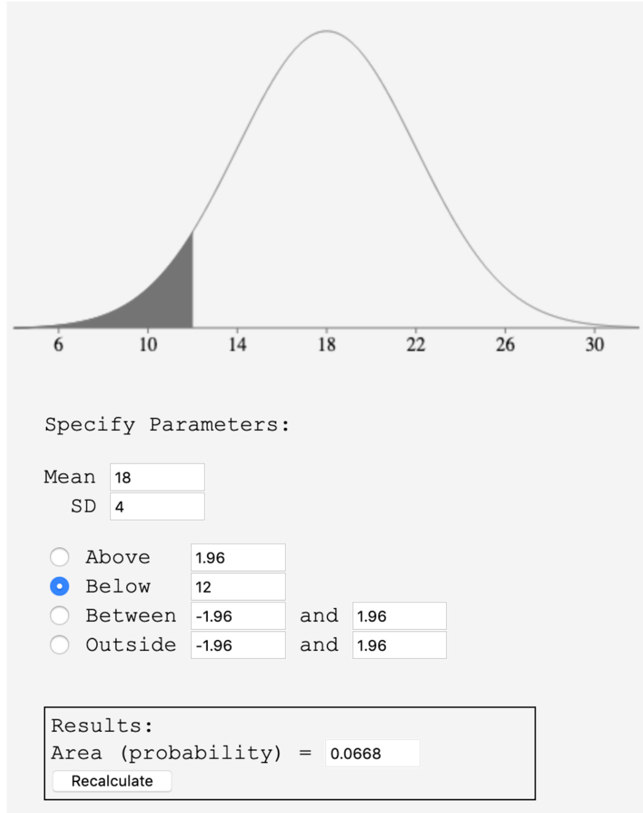
b. What percentage of students earned scores above 27 points?

c. What percentage of students earned scores between 21 and 25 points?

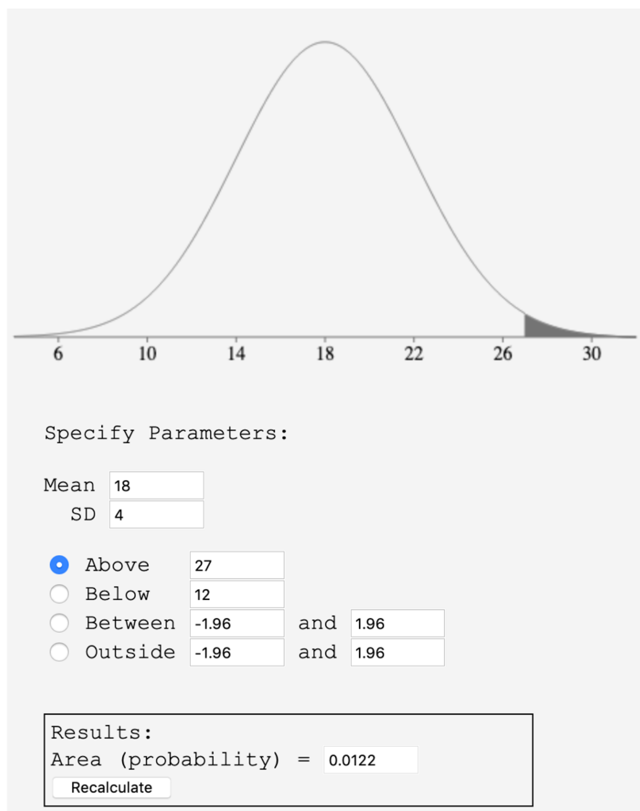
d. What percentage of students earned scores less than 11 or more than 23?

Solution

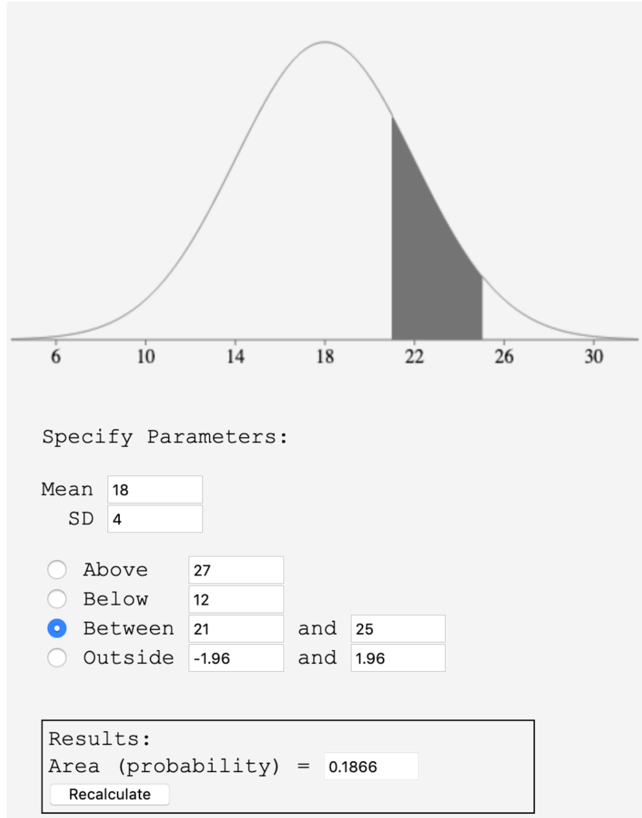
a. First, using OnlineStatbook, we we will enter the mean of 18 points and the standard deviation of 4 points. Then we select Below and enter 12. This gives us a result of 0.0668 as shown in the image or 6.68%. To use a spreadsheet, we enter `=NORM.DIST(12, 18, 4, 1)`. The last value is always a 1 to indicate that we want a cumulative probability or area. This also gives us a result of 0.0668 or 6.68%.



b. To use OnlineStatbook, we we will keep our mean of 18 points and standard deviation of 4 points. Then we select Above and enter 27. This gives us a result of 0.0122 as shown in the image or 1.22%. To use a spreadsheet, NORM.DIST gives us the area to the left so if we want the area to the right we need to find the complement or subtract from 1. We enter `=1-NORM.DIST(27, 18, 4, 1)`. This also gives us a result of 0.0122 or 1.22%.

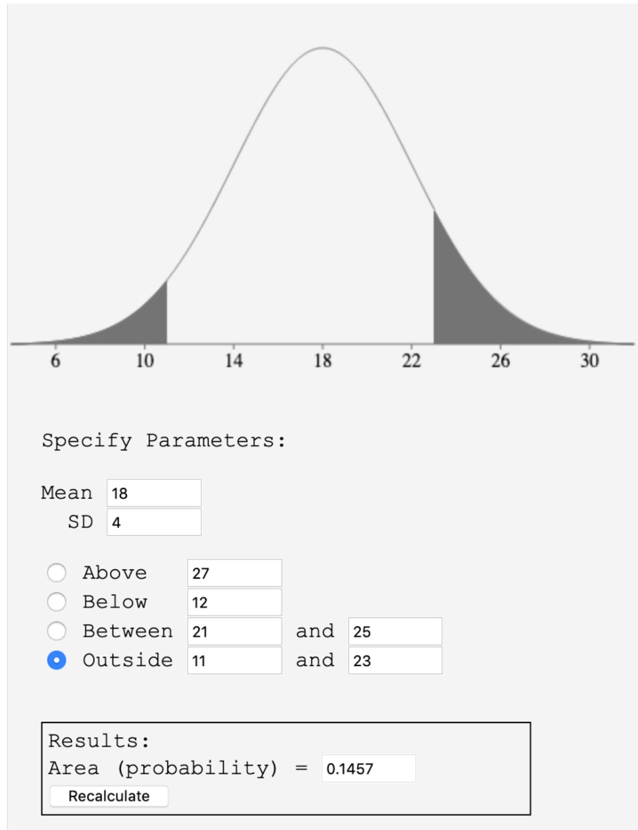


c. To use OnlineStatbook, we we will keep our mean of 18 points and standard deviation of 4 points. Then we select Between and enter 21 and 25. This gives us a result of 0.1866 as shown in the image or 18.66%.



To use a spreadsheet, we need to use NORM.DIST twice. We can find the area to the left of the upper value and then subtract the area to the left of the lower value. That will give us the area between the two values. We enter `=NORM.DIST(25, 18, 4, 1)-NORM.DIST(21, 18, 4, 1)`. This also gives us a result of 0.1866 or 18.66%.

d. To use OnlineStatbook, we we will keep our mean of 18 points and standard deviation of 4 points. Then we select Outside and enter 11 and 23. This gives us a result of 0.1457 as shown in the image or 14.57%.



To use a spreadsheet, we will find the left and right probabilities and add them together. We enter `=NORM.DIST(11, 18, 4, 1)+1-NORM.DIST(23, 18, 4, 1)`. This also gives us a result of 0.1457 or 14.57%.

□

Now you can this try this example.

Example 3.6.6 Continuing with the black lab dogs whose weights are approximately normally distributed with a mean of 78 pounds and a standard deviation 7 pounds, use technology to calculate the following probabilities:

a. The probability that a randomly chosen black lab weighs at least 95 pounds.

b. The probability that a randomly chosen black lab weighs less than 72 pounds.

c. The probability that a black lab weighs less than 65 pounds or greater than 86 pounds.

d. The probability that a black lab weighs between 66 and 75.

Solution.

a. Using OnlineStatbook and/or `=1-NORM.DIST(95, 78, 7, 1)` we get $P(X \geq 95) = 0.0076$ or 0.76%.

b. Using OnlineStatbook and/or `=NORM.DIST(72, 78, 7, 1)` we get $P(X < 72) = 0.1957$ or 19.57%.

c. Using OnlineStatbook and/or `=NORM.DIST(65, 78, 7, 1)+1-NORM.DIST(86, 78, 7, 1)` we get $P(X < 65) + P(X > 86) = 0.1582$ or 15.82%.

d. Using OnlineStatbook and/or `=NORM.DIST(75, 78, 7, 1)-NORM.DIST(66, 78, 7, 1)` we get $P(66 \leq X \leq 75) = 0.2909$ or 29.09%.

□

3.6.6 The Standard Normal Distribution and Z-Scores

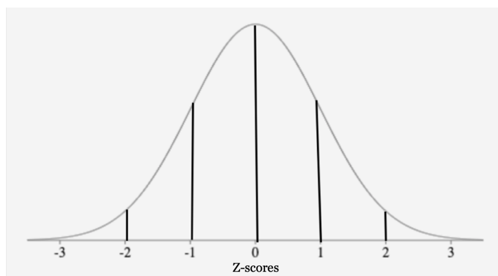
Earlier in the chapter we learned how to calculate a ***Z-score***, which is a way to determine how usual or unusual a data value is, or compare values from two different distributions. To calculate the Z-score for a data value, we calculate its distance from the mean. Then we divide that distance by the standard deviation to see how many standard deviations that is. Thus, the ***Z-score*** of a data value is the number of standard deviations it is away from the mean.

Z-score.

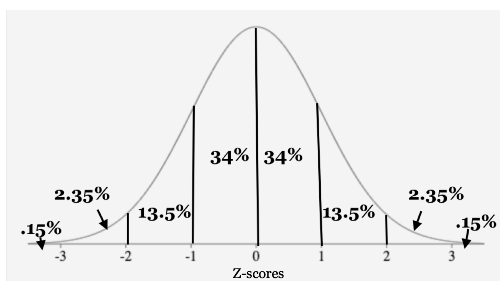
$$Z = \frac{\text{data value } - \text{ mean}}{\text{standard deviation}}$$

Be sure to calculate the difference first, then divide.

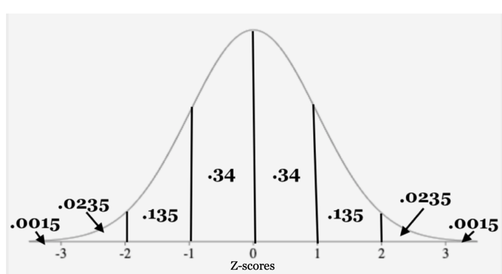
If a data value is equal to the mean it will have a Z-score of 0. If a data value is one standard deviation above the mean, it will have a Z-score of 1. We can make a Normal distribution of Z-scores and it will have a mean of 0 and a standard deviation of 1. This is called the ***Standard Normal distribution***, shown below.



Applying the Empirical Rule in percentage form to the Standard Normal curve looks like this.



Sometimes we may want to use the decimal form of the numbers instead.



Applying the Empirical Rule to the Standard Normal distribution, we know that 68% of all Z-scores will be between -1 and 1, 95% of all Z-scores will be between -2 and 2 and 99.7% of all Z-scores will be between -3 and 3. A Z-score below -3 or above 3 is possible, but is very unlikely.

Example 3.6.7 Let's say a student scored 27 points on the math quiz where the scores were approximately normally distributed with a mean of 18 points and a standard deviation of 4 points. We know from our earlier drawing that they scored between 2 and 3 standard deviations above the mean. Let's calculate the Z-score to determine exactly how many standard deviations from the mean they scored.

First we take the data value and subtract the mean of 18, then we divide by the standard deviation of 4 points:

$$Z = \frac{27 - 18}{4} = \frac{9}{4} = 2.25 \text{ standard deviations.}$$

The Z-score is positive so which confirms that they scored above the mean. □

Example 3.6.8 Now you can calculate the Z-score for a student who scored 15 on the math quiz.

Solution. First we take the data value and subtract the mean of 18, then we divide by the standard deviation of 4 points:

$$Z = \frac{15 - 18}{4} = \frac{(-3)}{4} = -0.75 \text{ standard deviations.}$$

This student scored 0.75 standard deviations below the mean. The Z-score is negative which means they scored below the mean. □

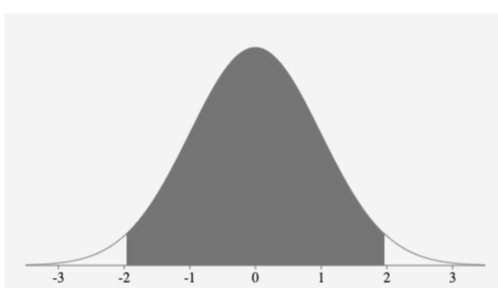
3.6.7 Confidence Intervals and Margin of Error

Now that we have learned more about the Normal distribution, we can come back to confidence intervals which we studied in section 3.1. When we take a sample, our goal is to infer our statistic to the larger population. Let's say we wanted to know the average height of all 10-year-old kids living in the United States. It would take too much time and money to measure them all so we take a random sample that is representative of the population of 10-year-olds.

If we measured 25 randomly selected 10-year-olds and calculated the sample mean of 54.5 inches, that statistic is called a ***point estimate***. It is a single value that we got from a sample. However, if we measured another random sample of 25 different kids we would likely get a different mean. That is called sampling variation or sampling error. To be more accurate in our estimation of the height of all 10-year olds we give a range of values called a ***confidence interval***.

To construct a confidence interval we need to know the point estimate, the standard error of the data and the desired confidence level. The ***standard error*** is an estimate for the standard deviation that we calculate from the sample data and the sample size. You can learn more details about sampling distributions, why they are Normally distributed and how to calculate the standard error in a statistics course.

The most common confidence level is the 95% confidence level. The 95% confidence interval represents the middle 95% of the Normal curve. From the Empirical rule we saw that approximately 95% of the data values are within 2 standard deviations of the mean. Using more exact methods, the middle 95% is 1.96 standard deviations from the mean.



Let's say that our point estimate for the height of 10-year-olds is 54.5 inches and the standard error is 1.75 inches. To estimate the middle 95% of values we multiply the standard error by 1.96 standard deviations.

$$\text{Point estimate } \pm 1.96 \times SE$$

$$54.5 \pm 1.96(1.75)$$

To find the lower boundary of the confidence interval we subtract

$$54.5 - 1.96(1.75) = 51.07 \text{ inches}$$

To find the upper boundary of the confidence interval add

$$54.5 + 1.96(1.75) = 57.93 \text{ inches.}$$

The 95% confidence interval is 51.07 to 57.93 inches or (51.07, 57.93) inches.

To interpret a confidence interval we say: *We are 95% confident that the true average height of all 10-year-olds is between 51.07 and 57.93 inches.* This means if repeated samples were taken and the 95% confidence interval was computed for each sample, about 95% of the intervals would contain the true population mean. About five percent of the intervals would not contain the population mean.

The standard error multiplied by the number of standard deviations for the confidence level is called the ***margin of error***. In this example we have

$$1.96 \times \text{SE}$$

$$1.96(1.75) = 3.43 \text{ inches.}$$

The margin of error is the range of values below and above your sample mean in the confidence interval.

Since the point estimate is our best guess for the value of the parameter, it makes sense to build the confidence interval around that value. The standard error, which is the measure of the uncertainty associated with the point estimate, provides a guide for how large we should make the confidence interval. When the sampling distribution of a point estimate can reasonably be modeled as Normal, the point estimate we observe will be within 1.96 standard errors of the true value of interest. We can be 95% confident this interval captures the true value.

The idea behind confidence intervals and margin of error is that any survey, study or poll will always be different from the population because samples vary. The confidence interval and margin of error reflect the sample size and the variation in the sample.

3.6.8 Exercises

1. Describe the shape of a Normal distribution.

2. In a Normal distribution, what percentage of observations will lie to the left of the mean?

3. If the diameters of trees in a forest follow a nearly Normal distribution has a mean of 35 inches and a standard deviation of 15 inches. What is the median width?

4. Two Normal distributions have the same means but different standard deviations. Distribution A has a standard deviation of 10 inches, and distribution B has a standard deviation of 15 inches. Which curve has a wider spread along the horizontal axis? Why?

Use the Empirical Rule to answer these questions.

5. About what percentage of the values from a Normal distribution fall within one standard deviation (left and right) of the mean?

6. About what percentage of the values from a Normal distribution fall within two standard deviations (left and right) of the mean?

7. About what percentage of the values from a Normal distribution fall within three standard deviations (left and right) of the mean?

8. About what percentage of the values from a Normal distribution fall outside of three standard deviations (left and right) of the mean?

9. About what percentage of the values from a Normal distribution fall between the first and second standard deviations from the mean (both sides)?

10. About what percentage of the values from a Normal distribution fall between the first and third standard deviations (both sides)?

11. About what percentage of values from a Normal distribution fall between the second and third standard deviations (on both sides)?

12. About what percentage of values from a Normal distribution fall outside the first standard deviation from the mean (on both sides)?

13. About what percent of values in a Normal distribution fall between the mean and one standard deviation below the mean?

14. About what percent of values in a Normal distribution fall between the mean and three standard deviations above the mean?

15. Suppose a Normal distribution has a mean of 6 inches and a standard deviation of 1.5 inches.

a. Draw and label the Normal distribution graph.

b. What is the range of data values that fall within one standard deviation of the mean?

c. What percentage of data fall between 3 and 10.5 inches?

d. What percentage of data fall below 1.5 inches?

16. Suppose a Normal distribution has a mean of 45 cm and a standard deviation of 10 cm.

a. Draw and label the Normal distribution graph.

b. What is the range of data values that fall within two standard deviations of the mean?

c. What percentage of the data fall between 15 and 55 cm?

d. What percentage of the data fall above 55 cm?

17. Suppose a Normal distribution has a mean of \$10 and a standard deviation of \$2.

a. Draw and label the Normal distribution graph.

b. What is the range of data values that fall within three standard deviations of the mean?

c. What percentage of data lie between \$6 and \$14?

d. What percentage of data lie above \$14?

18. Suppose the variable X is normally distributed with a mean of 15 miles and a standard deviation of 3 miles.

a. Draw and label the Normal distribution graph.

b. What is the range of data values that falls within one standard deviation of the mean?

c. What percentage of the data fall between 9 and 18 miles?

d. What percentage of the data fall above 18 or below 9 miles?

Use technology to answer these questions.

19. Suppose a Normal distribution has a mean of 15.5 ounces and a standard deviation of 4.2 ounces.

a. Draw and label the Normal distribution graph.

b. What percentage of the data values lie above 18.6 ounces?

c. What percentage of data lie between 9 and 20.2 ounces?

d. What percentage of data lie below 13.7 ounces?

20. Suppose a Normal distribution has a mean of 26.1 grams and a standard deviation of 6.5 grams.

a. Draw and label the Normal distribution graph.

b. What percentage of the data values fall above 32.6 grams?

c. What percentage of data is below 15 grams or greater than 36.7 grams?

d. What percentage of the data is less than or equal to 20.8 grams?

21. Suppose the variable X is normally distributed with a mean of 85 km and a standard deviation of 7 km.

a. Draw and label the Normal distribution graph.

b. Find $P(X \leq 82)$.

c. Find $P(76 \leq X \leq 90)$.

d. Find $P(X \geq 100)$.

22. Suppose the variable X is normally distributed with a mean of \$7 and a standard deviation of \$1.30.

a. Draw and label the Normal distribution graph.

b. Find $P(X \geq 8)$.

c. Find $P(X \leq 4.1)$ or $P(X \geq 7.8)$.

d. Find $P(X \leq 7.3)$.

23. What does a z-score measure?

24. Consider the Standard Normal distribution. The mean is always ________ and the standard deviation is always _________.

25. For the distribution in problem 15, calculate the Z-score for a data value of 6.2 inches.

26. For the distribution in problem 16, calculate the Z-score for a data value of 32 cm.

27. For the distribution in problem 17, calculate the Z-score for a data value of \$5.

28. For the distribution in problem 18, calculate the Z-score for a data value of 19 miles.

29. A random sample of 45 people who carry a purse found that they had an average of \$2.35 in change in the bottom of their purse. The margin of error was \$0.15. Calculate the 95% confidence interval and interpret the results.

30. Scores on a certain quiz are normally distributed. In sample of 25 students the mean score was 14 points with a standard error estimate of 2 points. Calculate the margin of error and the 95% confidence intervals. Interpret Confidence interval.

31. A researcher works on a study and found the sample mean to be 84.5 cm and the standard error of estimate to be 0.11 cm. What is the margin of error for the 95% confidence level? What is the 95% confidence interval for the true mean? Interpret the results.

32. A researcher works on a study and found that the sample mean to be \$35.4 and the standard error of estimate to be \$0.75. What is the margin of error for the 95% confidence level? What is the 95% confidence interval for the true mean? Interpret the results.

Chapter 4

Probability

Dr. Hilda Geiringer and Probability.

Dr. Hilda Geiringer[1] (1893 - 1973) was born in Vienna, Austria, to a Jewish family. When she was in high school she showed great mathematical ability and while many women were discouraged from academics, her parents supported her so she could attend the University of Vienna. She earned her Ph.D. in mathematics in 1917. In 1921, she moved to Berlin, Germany, where she was a research assistant to Richard von Mises in the Institute of Applied Mathematics (Robertson & O'Connor, 2020).
Dr. Geiringer was the first woman to teach applied mathematics at a German university (Manning, 2019) and in 1933, she was nominated for an assistant professor position. In the same year, however, the Nazi party took over in Germany and she was dismissed as a result of Nazi anti-Jewish legislation (Freidenreich, 2020).

After some time in Brussels, Belgium, she fled again to Turkey who welcomed approximately 200 German scholars. She had to learn Turkish to teach and became a professor with a five-year contract (McNeill, 2019). It was there that she thrived, publishing 18 articles and a calculus textbook. She conducted innovative research in probability theory and Mandelian genetics by configuring recursive equations to study the distribution of genotypes and blood types (McNeill, 2019).

In 1939, Turkey began replacing Jewish refugees at the university and fearing for her safety again, Dr. Geiringer had another challenge in moving to the United States. In the U.S. she faced open discrimination towards women in mathematics and she could only teach at womens colleges. She accepted an unpaid position at Bryn Mawr College to get a visa to the U.S. (McNeill, 2019) and then settled at Wheaton College in Norton, Massachusetts.

While Wheaton was not a research institution, she never gave up her research and she published more than 80 articles, reviews and books on pure and applied mathematics (Manning, 2019). In 1953 she wrote, "I have to work scientifically, besides my college work. This is a necessity for me; I never stopped it since my student days, it is the deepest need of my life."

Dr. Geiringer is also known for the Geiringer equations that simplify the process of calculating the deformation of metal in slip-line field theory. In 1956 she was elected professor emeritus by the University of Berlin with a full salary and she retired from Wheaton in 1959 (Robertson & O'Connor, 2020).

[1]Photo: "Hilda Geiringer Portrait Photo"(cropped) from the Marion B. Gebbie Archives and Special Collections, Wheaton College is licensed under CC BY-SA 4.0.

4.1 Contingency Tables

Objectives: Section 4.1 Contingency Tables

Students will be able to:

- Relate Venn diagrams and contingency tables
- Calculate percentages from a contingency table
- Calculate "and" empirical probabilities
- Calculate "or" empirical probabilities
- Calculate conditional probabilities
- Determine whether two characteristics are independent

When we looked at categorical data in the previous chapter, it was related to a single variable, or characteristic of interest, such as favorite movie or car color. To illustrate the data, we made a frequency table and used it to create a pie chart or bar chart. But what if we want to illustrate the relationship between two categorical variables? To do this, we can use a contingency table.

4.1.1 Contingency Tables

A ***contingency table*** summarizes all the possible combinations for two categorical variables. Each value in the table represents the number of times a particular combination of outcomes occurs. For example, suppose we randomly select 250 households from the greater Portland area and ask whether they have a cat and whether they have a dog. In this case, "have a cat" and "have a dog" are the two variables, and each variable has two categories: Yes and No. To create the contingency table, we make columns for the categories of one variable, and rows for the categories of the other variable. We also add a row and column for the subtotals of each category. Each cell of the resulting table contains the number of outcomes having the characteristics of the intersecting row and column categories. For our dog and cat example, the table would look like this:

	Dog	***No Dog***	***Total***
Cat	*Yes Cat and Yes Dog*	*Yes Cat and No Dog*	*Yes Cat Total*
No Cat	*No Cat and Yes Dog*	*No Cat and No Dog*	*No Cat Total*
Total	*Yes Dog Total*	*No Dog Total*	*Grand total*

Suppose that of the 250 households surveyed, 180 said they have a cat, 95 said they have a dog, and 52 said they have both a cat and a dog. We can use this information to fill in the cells of the table.

	Dog	***No Dog***	***Total***
Cat	52		180
No Cat			
Total	95		250

The first cell we can fill in is the ***grand total***, which is the total number of subjects in the study. In this case, there are 250 households participating in the survey. The next two cells we can fill in are the total number of households that have a cat, 180, and the total number of households that have a dog, 95. The final cell we can fill in from the given information is the intersection of the having a dog column and a having a cat row, which is 52 households.

Since each row and column must sum to their totals, we can use subtraction to find the missing numbers as shown below.

	Dog	***No Dog***	***Total***
Cat	52	$180 - 52 = 128$	180
No Cat	$95 - 52 = 43$	$155 - 128 = 27$ or $70 - 43 = 27$	$250 - 180 = 70$
Total	95	$250 - 95 = 155$	250

Now that we have our contingency table completed, notice that the numbers in the central four cells add to the grand total as shown in the table on the left. The total row and the total column also add to the grand total as shown in the right table.

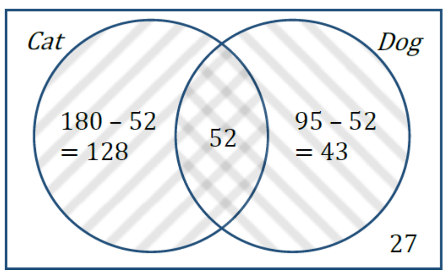
	Dog	***No Dog***	***Total***
Cat	52	128	180
No Cat	43	27	70
Total	95	155	250

	Dog	***No Dog***	***Total***
Cat	52	128	180
No Cat	43	27	70
Total	95	155	250

4.1.2 Contingency Tables and Venn Diagrams

If the subtractions we just did seem familiar, they should! This is very similar to what we did for reporting data with a Venn diagram. The Venn diagram for this data is shown below. We also subtracted the intersection from the total of the cat and dog owners to find numbers in the crescent regions.

Notice that the numbers in the four regions of the Venn diagram are the same as the four cells in the center of the contingency table and add to the grand total.



4.1.3 "And" Statements

Now we can use the contingency table or the Venn diagram to determine the percentage of households that meet certain conditions. For instance, what percent of those surveyed own a cat ***and*** do not own a dog? In the Venn diagram, this is 128 households in the cat only region.

In the contingency table we see the 128 households at the intersection of the row of households who own a cat and the column of households who do not own a dog. As a percentage, the total number of households surveyed, is $\frac{128}{250} = 0.512$ or 51.2% that have a cat and no dog.

	Dog	***No Dog***	***Total***
Cat	52	128	180
No Cat	43	27	70
Total	95	155	250

4.1.4 "Or" Statements

How about the percentage of households surveyed that have a cat ***or*** a dog? We know from Venn diagrams that the inclusive *or* includes the number of households who own a cat only, a dog only, and both a cat and a dog, or $128 + 52 + 43 = 223$ households. As a percentage of the total surveyed, we get $\frac{223}{250} = 0.892$ or 89.2% of households in the sample have a dog ***or*** a cat (or both).

We can get the same answer from the contingency table. by adding the cells for households who have a cat and not a dog, a dog and not a cat, and the households that have both a cat and a dog. This also gives us 223 households.

There is another way to calculate an *or* statements from a contingency table. We could add the row and column totals for having a cat and having a dog, but then we have counted the 52 households in the intersection twice. We can subtract that number to get $180 + 95 - 52 = 223$ households with a dog or a cat, which we know is 89.2% of those surveyed.

	Dog	***No Dog***	***Total***
Cat	52	128	180
No Cat	43	27	70
Total	95	155	250

4.1.5 Conditional Statements

Another question we can answer using a contingency table is what percentage of dog owning households also own a cat? In this case the group that we are interested in isn't every household surveyed (the grand total), but just those households that own a dog.

	Dog	***No Dog***	***Total***
Cat	52	128	180
No Cat	43	27	70
Total	95	155	250

We call this a ***conditional*** statement because we are only considering the households with a certain condition. If we focus on the column representing the households that own a dog, we see that there is a total of 95 households with a dog, and that 52 of those 95 households also have a cat. Therefore, $\frac{52}{95} \approx 0.547$ or approximately 54.7% of the households with a dog also have a cat. Another way to phrase this conditional statement is, "What percent of households have a cat ***given*** they have a dog." You will see the word given quite a bit in this chapter and that makes the denominator change. It is also possible to find this conditional percentage using the Venn diagram by taking the number in the intersection and dividing it by the total in the whole dog circle.

4.1.6 Contingency Tables with More Than Two Categories

When there are only two categories for each variable, like yes/no questions, Venn diagrams and contingency tables provide basically the same information and can be used interchangeably. A Venn diagram works well for yes/no variables since a subject is either inside the circle (has the characteristic) or outside the circle (does not have the characteristic). If we have more than two possibilities for any of the variables, though, we cannot use a Venn diagram. We can use a contingency table, though. Here is an example where one variable has four categories and the other has three categories.

Example 4.1.1 910 randomly sampled registered voters from Tampa, FL were asked if they thought workers who have illegally entered the US should (i) be allowed to keep their jobs and apply for US citizenship, (ii) be allowed to keep their jobs as temporary guest workers but not be allowed to apply for US citizenship, or (iii) lose their jobs and have to leave the country. Not sure was also an option (iv). The results of the survey by political ideology are shown below[1]. Use the contingency table to answer the questions.

	Conservative	***Moderate***	***Liberal***	***Total***
(i) Apply for citizenship	57	120	101	278
(ii) Guest worker	121	113	28	262
(iii) Leave the country	179	126	45	350
(iv) Not sure	15	4	1	20
Total	372	363	175	910

a. What percent of the sampled Tampa, Fl voters identified themselves as conservatives?

b. What percent of the sampled voters are in favor of the citizenship option?

c. What percent of the sampled voters identify themselves as conservatives ***and*** are in favor of the citizenship option?

d. What percent of the sampled voters identify themselves as liberal ***or*** are in favor of the leaving the country option?

e. What percent of the sampled voters who identify as conservatives are also in favor of the citizenship option? What percent of moderate and liberal voters share this view?

Solution.

1. To answer this question, we find the conservative column and look to the bottom cell for the total number of conservative voters and divide that by the total number of voters surveyed. This gives us $\frac{372}{910} \approx 0.409$ or approximately 41% of the Tampa, Fl voters who identify as conservative.

2. For this question we find the apply for citizenship row, look across to find the total, and divide this by the total number of voters surveyed. We get $\frac{278}{910} \approx 0.305$ or approximately 31% of these voters are in favor of the citizenship option.

3. For this question we are looking for the cell that is the intersection of those who identify as conservative and those who are in favor of the citizen option. This cell has 57 voters, so we divide that by the total number of voters. This gives us or $\frac{57}{910} \approx 0.063$ or approximately 6.3% of these voters identify as conservatives and are in favor of the citizenship option.

4. The *or* in this question is inclusive, so we need to determine the number of voters who identify as liberal, who are in favor of the leaving the country option, or both.

	Conservative	***Moderate***	***Liberal***	***Total***
(i) Apply for citizenship	57	120	101	278
(ii) Guest worker	121	113	28	262
(iii) Leave the country	179	126	45	350
(iv) Not sure	15	4	1	20
Total	372	363	175	910

In terms of the individual cells, the number of voters who have the specified characteristics is the sum $179 + 126 + 101 + 28 + 45 + 1 = 480$, which we can divide by the total number of voters surveyed to get the percent. So, we have $\frac{480}{910} \approx 0.527$ or approximately 53% of the voters identify as liberal or are in favor of the leave the country option.

Another way to calculate this is to add the total number who identify as liberal (175 voters) and the total number who are in favor of the leave the country option (350 voters), then subtract the double counted cell (45 voters) who are liberal and in favor of the leave the country option: $175 + 350 - 45 = 480$

5. As we saw before, these are conditional statements. For the first part of this question, we want to focus just on those voters who identify as conservatives, and from among that group determine the percent in

favor of the citizenship option. We calculate that $\frac{57}{372} \approx 0.153$ or approximately 15% of conservative voters are in favor of the citizenship option.

For the second part, we want to focus on just those voters who identify as moderate, and from among that group determine the percent in favor of the citizenship option. Then we have $\frac{120}{363} \approx 0.33$ or approximately 33% of moderate voters are in favor of the citizen option.

Finally, we want to focus on just those voters who identify as liberal, and from among that group determine the percent in favor of the citizenship option. We calculate $\frac{101}{175} \approx 0.58$ or approximately 58% of liberal voters are in favor of the citizenship option. Looking at these three percentages, it is clear that support of the citizenship option ***depends*** on political ideology. If support of the citizenship option were the same across political ideologies, then we would say that favoring the citizenship option and political ideology were ***independent*** of each other.

□

4.1.7 Empirical Probability

If our sample is representative of the population, then we can also interpret a percentage we calculate from a contingency table as a ***probability***, or the likelihood that something will happen. Since a contingency table is constructed from data collected through sampling or an experiment, we call it an ***empirical*** or ***experimental*** probability. This is different from a ***theoretical*** probability which we will look at in the next section.

4.1.8 Finding Empirical Probabilities with a Contingency Table

Suppose that 60% of students in our class have a summer birthday (June, July, or August). Now suppose everyone's name and birth month are written on slips of paper and thrown into a bag. If we pull a slip of paper out of the bag at random, what is the probability that the selected student has a summer birthday? If you think there should be a 60% chance, you are right! The relative frequency of the characteristic of interest will be equal to its empirical probability. To write this as a probability statement, it would look like

$$\text{P(summer birthday)} = 60\%$$

Probability is a function named P, and the function is applied to what follows in the parentheses. Let's look at another example where we write probability statements and find empirical probabilities.

Example 4.1.2 A survey of licensed drivers asked whether they had received a speeding ticket in the last year and whether their car is red. The results of the survey are shown in the contingency table to the right.

	Speeding Ticket	***No Speeding Ticket***	***Total***
Red Car	15	135	150
Not Red Car	45	470	515
Total	60	605	665

Find the probability that a randomly selected survey participant:

a. has a red car.

b. has had a speeding ticket in the last year.

c. has a red car and has not had a speeding ticket in the last year.

[1]SurveyUSA, News Poll #18927, data collected Jan 27-29, 2012. Example adapted from Open Intro: Advanced High School Statistics, by Diez et al, used under CC-BY-SA 3.0.

d. has a red car or has had a speeding ticket in the last year.

e. has had a speeding ticket in the last year given they have a red car.

f. who has received a speeding ticket in the last year also has a red car.

g. What do the answers to b and e suggest about the relationship between owning a red car and getting a speeding ticket?

Solution.

a.

	Speeding Ticket	***No Speeding Ticket***	***Total***
Red Car	15	135	150
Not Red Car	45	470	515
Total	60	605	665

To find $P(\text{red car})$, we divide the number of participants who own a red car by the total number of people surveyed:$P(\text{red car}) = \frac{150}{665} \approx 0.226$ or 22.6%.

b.

	Speeding Ticket	***No Speeding Ticket***	***Total***
Red Car	15	135	150
Not Red Car	45	470	515
Total	60	605	665

$P(\text{speeding ticket})$, we divide the number of participants who got a speeding ticket in the last year by the total number of people surveyed: $P(\text{speeding ticket}) = \frac{60}{665} \approx 0.09$ or 9%.

c.

	Speeding Ticket	***No Speeding Ticket***	***Total***
Red Car	15	135	150
Not Red Car	45	470	515
Total	60	605	665

To find $P(\text{red and no ticket})$, we find the intersection of the red car category and the no ticket category and divide by the total number of participants: $P(\text{red and no ticket}) = \frac{135}{665} \approx 0.203$ or 20.3%

d.

	Speeding Ticket	***No Speeding Ticket***	***Total***
Red Car	15	135	150
Not Red Car	45	470	515
Total	60	605	665

To find $P(\text{red or ticket})$, we need to add those who drive a red car and did not have a speeding ticket (just red), those who had a speeding ticket and do not drive a red car (just ticket) and those who drive a red car and had a speeding ticket (both), and divide by the total number of participants:

$$P(\text{red and no ticket}) = \frac{135 + 45 + 15}{665} = \frac{195}{665} \approx 0.293 \text{ or } 29.3\%$$

Recall from our earlier discussion that we could also calculate the or probability as:

$$\begin{aligned} P(\text{red and no ticket}) &= P(\text{red}) + P(\text{speeding ticket}) - P(\text{red and speeding ticket}) \\ &= \frac{150}{665} + \frac{60}{665} - \frac{15}{665} \\ &= \frac{195}{665} \end{aligned}$$

which gives us the same answer as counting the individual cells.

e.

	Speeding Ticket	***No Speeding Ticket***	***Total***
Red Car	15	135	150
Not Red Car	45	470	515
Total	60	605	665

The probability P(speeding ticket given red car) is a conditional probability as we have seen before since it is conditional on the given characteristic occurring. In this problem, the given characteristic is owning a red car, so we isolate our attention to just the row of 150 red car owners and see how many have had a speeding ticket in the last year. Looking at the table, we see that there were 15 red car owners who had a speeding ticket in the last year, so we calculate:

$$P(\text{speeding ticket given red car}) = \frac{15}{150} = 0.10 \text{ or } 10\%$$

f.

	Speeding Ticket	***No Speeding Ticket***	***Total***
Red Car	15	135	150
Not Red Car	45	470	515
Total	60	605	665

This question is also asking for a conditional probability, P(red car given speeding ticket), but it is phrased more like we would say it. In this case the given characteristic is that the person has received a speeding ticket, so we will isolate our attention to just the speeding ticket column. Among the 60 people who had a speeding ticket in the last year, we see that 15 also drove a red car. Now we can calculate the probability:

$$P(\text{red car given speeding ticket}) = \frac{15}{60} = 0.25 \text{ or } 25\%$$

Notice that compared with part e, when we change the conditional characteristic, we change the denominator of the fraction.

g. In part b, we determined that there was a 9% chance of randomly selecting a participant who had received a speeding ticket in the last year. However, in part e we found that there was a 25% chance of receiving a ticket in the last year if the person had a red car. This seems to suggest that there is a higher likelihood of getting a speeding ticket if you own a red car. This means that getting a speeding ticket is **dependent** on whether the person drives a red car, since that increases the probability of getting a ticket. We cannot say, however, whether driving a red car makes you speed or whether people who tend to drive faster buy red cars.

□

4.1.9 Conditional Probabilities

We have mentioned conditional probabilities, which we find by isolating our attention to the given row or column. Here is another example of finding conditional probabilities.

Example 4.1.3 A home pregnancy test was given to a sample of 93 cisgender women, and their pregnancy was then verified by a blood test. The contingency table below shows the home pregnancy test and whether or not they were actually pregnant as determined by the blood test. Find the probability that a randomly selected woman in the sample

a. was not pregnant given the home test was positive.

b. had a positive home pregnancy test given they were not pregnant.

	Positive Test	***Negative Test***	***Total***
Pregnant	70	4	74
Not Pregnant	5	14	19
Total	75	18	93

Solution. Here are the solutions:

a. Since we are given the home test result was positive, we are limited to the 75 women in the positive test column, of which 5 were not pregnant. This gives:

	Positive Test	***Negative Test***	***Total***
Pregnant	70	4	74
Not Pregnant	5	14	19
Total	75	18	93

$$P(\text{not pregnant given positve test}) = \frac{5}{75} \approx 0.067 \text{ or } 6.7\%$$

b. Since we are given the woman is not pregnant, we are limited to the 19 women in the not pregnant row, of which 5 had a positive test. This gives:

	Positive Test	***Negative Test***	***Total***
Pregnant	70	4	74
Not Pregnant	5	14	19
Total	75	18	93

$$P(\text{positive test given not pregnant}) = \frac{5}{19} \approx 0.263 \text{ or } 26.3\%$$

This result is referred to as a false positive: A positive test result when the woman is not actually pregnant.

□

In this section we have learned about empirical probability. In the next section we will discuss another kind of probability that you may be familiar with – theoretical probability.

4.1.10 Exercises

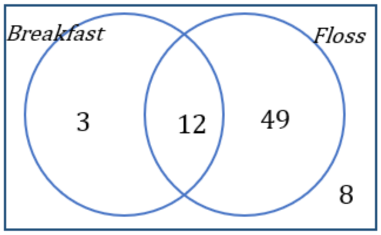
1. A recent survey asked a random sample of PCC students if they are currently experiencing food insecurity and if they are currently experiencing housing insecurity. Fill in the missing entries of the contingency table below.

	Food Insecure	Not Food Insecure	Total
Housing Insecure		60	
Not Housing Insecure		460	760
Total	680		

2. A recent survey asked a random sample of PCC students if they have purchased food from the cafeteria in the last week, and if they purchased their textbooks through the bookstore. Fill in the missing entries of the contingency table below.

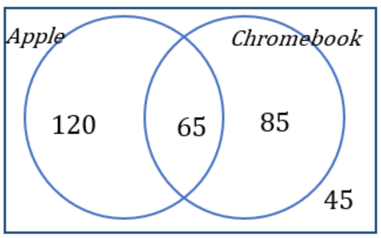
	Bookstore	No Bookstore	Total
Cafeteria			375
No Cafeteria		135	
Total	630		850

3. A recent survey asked PCC students if they regularly eat breakfast and if they regularly floss their teeth Use the completed Venn Diagram to fill in the corresponding contingency table.



	Breakfast	No Breakfast	Total
Floss			
No Floss			
Total			

4. A recent survey asked PCC students if they used an Apple phone, and if the regularly used a Chromebook outside of school. Use the completed Venn Diagram to fill in the corresponding contingency table.



	Chromebook	No Chromebook	Total
Apple			
No Apple			
Total			

5. Use the following information to complete the contingency table:

- P(A and B) = 10/75
- P(A) = 40/75
- P(not B) = 45/75

	A	Not A	Total
B			
Not B			
Total			

6. Use the following information to complete the contingency table:

- P(A given B) = 30/80
- P(Not A and Not B) = 10/120

	A	Not A	Total
B			
Not B			
Total			

7. A professor gave a test to students in a morning class and the same test to the afternoon class. The grades are summarized below.

	A	B	C	Total
Morning Class	8	18	13	39
Afternoon Class	10	4	12	26
Total	18	22	25	65

If one student was chosen at random:

a. What is the probability they were in the morning class?

b. What is the probability they earned a C?

c. What is the probability that they earned an A and they were in the afternoon class?

d. What is the probability that they earned an A given they were in the morning class?

e. What is the probability that they were in the morning class or they earned a B?

8. A professor surveyed students in her morning and afternoon Math 105 class, and asked what their class standing was. The class standings are summarized below:

	Freshman	Sophomore	Junior	Senor	Total
Morning Class	12	5	7	8	32
Afternoon Class	5	13	8	2	28
Total	17	18	15	10	60

If one student was chosen at random:

a. What is the probability they were in the morning class?

b. What is the probability they were a Freshman?

c. What is the probability that they were a Senior and they were in the afternoon class?

d. What is the probability that they were a Sophomore given they were in the morning class?

e. What is the probability that they were in the morning class or they were a Junior?

9. The contingency table below shows the number of credit cards owned by a group of individuals below the age of 35 and above the age of 35.

	Zero	One	Two or more	Total
Between the ages of 18-35	9	5	19	33
Over age 35	18	10	20	48
Total	27	15	39	81

If one person was chosen at random:

a. What is the probability they had no credit cards?

b. What is the probability they had one credit card?

c. What is the probability they had no credit cards and is over 35?

d. What is the probability they are between the ages of 18 and 35, or have zero credit cards?

e. What is the probability they had no credit cards given that they are between the ages of 18 and 35?

f. What is the probability they have no credit cards given that they are over age 35?

g. Does it appear that having no credit cards depends on age? Or are they independent? Use probability to support your claim.

10. The following contingency table provides data from a sample of 6,224 individuals who were exposed to smallpox in Boston.[2]

	Inoculated	Not Inoculated	Total
Lived	238	5136	5374
Died	6	844	850
Total	244	5980	6224

a. What is the probability that a person was inoculated?

b. What is the probability that a person lived?

c. What is the probability that a person died or was inoculated?

d. What is the probability that a person died given they were inoculated?

e. What is the probability that a person died given they were not inoculated?

f. Does it appear that survival depended on if a person were inoculated? Or are they independent? Use probability to support your claim.

11. The contingency table below shows the survival data for the passengers of the Titanic.

	First	Second	Third	Crew	Total
Survive	203	118	178	212	711
Not Survive	122	167	528	673	1490
Total	325	285	706	885	2201

a. What is the probability that a passenger did not survive?

b. What is the probability that a passenger was crew?

c. What is the probability that a passenger was first class and did not survive?

d. What is the probability that a passenger did not survive or was crew?

e. What is the probability that a passenger survived given they were first class?

f. What is the probability that a passenger survived given they were second class?

g. What is the probability that a passenger survived given they were third class?

h. Does it appear that survival depended on the passenger's class? Or are they independent? Use probability to support your claim.

12. The following table shows the utility patents granted for a specific year.

	Corporation	Government	Individual	Total
United States	45%	2%	8%	55%
Foreign	41%	1%	3%	45%
Total	86%	11%	3%	100%

a. What is the probability that a patent is foreign and from the government?

b. What is the probability that a patent is from the U.S. and from a corporation?

c. What is the probability that a patent is foreign or from the government?

d. What is the probability that a patent is from the U.S. given it is from an individual?

e. What is the probability that a patent is foreign given it is from the government?

[2] Data taken from Mostly Harmless Probability & Statistics by Rachel Webb

13. There is a 15% chance that a shopper entering a computer store will purchase a computer, a 25% chance they will purchase a game/software, and there is a 10% chance they will purchase both a computer and a game/software.

a. Create a contingency table for the information.

	Game/Software	No Game/Software	Total
Computer			
No Computer			
Total			

b. What is the probability that a shopper will not purchase a computer and will not purchase a game/software?

c. What is the probability that a shopper will purchase a computer or purchase a game/software?

d. What is the probability that a shopper will purchase a game/software given they have purchased a computer?

e. What is the probability that a shopper will purchase a game/software given they did not purchase a computer?

f. Does it appear that purchasing a game/software depends on whether the shopper purchased a computer? Or are they independent? Use probability to support your claim.

14. A fitness center coach kept track over the last year of whether members stretched before they exercised, and whether or not they sustained an injury. Among the 400 members, 322 stretched before they exercised, 327 did not sustain an injury, and 270 both stretched and did not sustain an injury.

a. Create a contingency table for the information.

	Injury	No Injury	Total
Stretched			
Not Stretched			
Total			

b. What is the probability that a member sustained an injury?

c. What is the probability that a member sustained an injury and did not stretch?

d. What is the probability that a member stretched or did not sustain an injury?

e. What is the probability that a member sustained an injury given they stretched?

f. What is the probability that a member sustained an injury given they did not stretch?

g. Does it appear that sustaining an injury depends on whether the member stretches before exercising? Or are they independent? Use probability to support your claim.

15. Among the 95 books on a bookshelf, 72 are fiction, 28 are hardcover, and 87 are fiction or hardcover.

a. Create a contingency table for the information.

	Hardcover	Paperback	Total
Fiction			
Nonfiction			
Total			

b. What is the probability that a book is non-fiction and paperback?

c. What is the probability that a book is fiction given it is hardcover?

16. After finishing the course, among the 32 students in a Math 105 class, 25 could successfully construct a contingency table, 27 passed the class, and 29 could successfully construct a contingency table or passed the class.

a. Create a contingency table for the information.

	Contingency Table	No Contingency Table	Total
Pass			
No Pass			
Total			

b. What is the probability that a student passed and could not successfully construct a contingency table?

c. What is the probability that a student passed given they could not successfully construct a contingency table?

4.2 Theoretical Probability

Objectives: Section 4.2 Theoretical Probability

Students will be able to:

- Write the sample space for theoretical probability situations
- Identify certain and impossible events
- Calculate the theoretical probability of a complement
- Determine the difference between empirical and theoretical probability
- Explain the Law of Large Numbers
- Identify independent and dependent events
- Calculate "and" theoretical probabilities
- Identify overlapping and disjoint sets
- Calculate "or" theoretical probabilities
- Calculate probability values for simple games

As we saw in the last section, the ***probability*** of a specified event is the chance or likelihood that it will occur. We calculated empirical or experimental probabilities using contingency tables. In this section, we will focus on ***theoretical*** probability and compare the two types.

4.2.1 Basic Probability Concepts

Let's begin with a brief introduction to some of the language and basic concepts of theoretical probability.

4.2.2 Experiment

If you roll a die, pick a card from a deck of playing cards, or randomly select a person and observe their hair color, you are conducting an ***experiment***.

4.2.3 Events and Outcomes

The result of an experiment is an ***outcome***, and a particular outcome, like rolling a five on a die, is called an ***event***. An event can be a ***simple event*** or combination of outcomes, called a ***compound event***.

4.2.4 Sample Space

The ***sample space*** is the set of all possible outcomes. For example, if we roll a six-sided die, the sample space S is the set $S = 1, 2, 3, 4, 5, 6$.

Example 4.2.1 If we roll an eight-sided die, describe the sample space and give at least two examples of simple events and compound events.

Solution. The sample space is the set of all possible outcomes, or equivalently, all simple events: $S = 1, 2, 3, 4, 5, 6, 7, 8$

Examples of simple events are rolling a 1, rolling a 5, rolling a 6, and so on. Examples of compound events include rolling an even number, rolling a 5 or a 3, and rolling a number that is at least 4. □

4.2.5 Equally Likely Outcomes

When the outcomes of an experiment are equally likely, we can calculate the probability of an event as the number of ways it can happen out of the total number of outcomes.

Theoretical Probability.

$$P(E) = \frac{\text{number of outcomes corresponding to the event E}}{\text{total number of equally-likely outcomes}}$$

We can write the result as a simplified fraction or as a decimal or percent.

Example 4.2.2 Write the sample space for the sum of two fair six-sided dice and determine whether the outcomes are equally likely.

Solution. The sample space for the sum of two fair six-sided dice is $S = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$. The different sums, however, are not equally likely. If we look at a table of the different possible outcomes when rolling two dice, we see that there are 36 possible combinations. We will summarize this in a table by listing each outcome in the sample space. Then we find the probabilities by counting the number of ways each sum can occur and dividing it by 36.

	1	2	3	4	5	6
1	$1+1=2$	$1+2=3$	$1+3=4$	$1+4=5$	$1+5=6$	$1+6=7$
2	$2+1=3$	$2+2=4$	$2+3=5$	$2+4=6$	$2+5=7$	$2+6=8$
3	$3+1=4$	$3+2=5$	$3+3=6$	$3+4=7$	$3+5=8$	$3+6=9$
4	$4+1=5$	$4+2=6$	$4+3=7$	$4+4=8$	$4+5=9$	$4+6=10$
5	$5+1=6$	$5+2=7$	$5+3=8$	$5+4=9$	$5+5=10$	$5+6=11$
6	$6+1=7$	$6+2=8$	$6+3=9$	$6+4=10$	$6+5=11$	$6+6=12$

Sum	Probability
2	$1/36$
3	$2/36$
4	$3/36$
5	$4/36$
6	$5/36$
7	$6/36$
8	$5/36$
9	$4/36$
10	$3/36$
11	$2/36$
12	$1/36$

From the probability table we can see that rolling a sum of 7 has the highest probability and rolling a 2 or a 12 have the lowest probabilities. □

Example 4.2.3 Suppose we roll a fair six-sided die. Calculate the probability of:

1. rolling a 6.
2. rolling a number that is at least 4.
3. rolling an even number.

4. rolling a 5 or a 3.

Solution. Recall that the sample space is $S = 1, 2, 3, 4, 5, 6$. Since each of the outcomes in the sample space is equally likely, we can find the probability of each event by counting the number of outcomes corresponding to the event and dividing by 6, the total number of equally likely outcomes.

1. There is only one way to roll a 6, so $P(\text{rolling a 6}) = \frac{1}{6}$ or approximately 16.7% There is a 16.7% chance of rolling a 6.
2. In probability we will often come across the phrases "at least" and "at most." ***At least*** means that value or greater. ***At most*** means that value or less. Since we are looking for the probability of rolling a number that is at least 4, we need the number of outcomes that are 4 or greater. There are 3 values that meet this condition: 4, 5, and 6. The probability is $P(\text{rolling a number that is at least 4}) = \frac{3}{6} = \frac{1}{2}$ or 50%
3. Half of the numbers on a die are even, so we calculate: $P(\text{rolling an even number}) = \frac{3}{6} = \frac{1}{2}$ or 50%
4. There are two ways to roll a 5 or a 3, so $P(\text{rolling a 5 or a 3}) = \frac{2}{6} = \frac{1}{3}$ or approxiamtely 33.3%. There is a 33.3% chance of rolling a 5 or a 3.

□

Example 4.2.4 Suppose you have a bag containing 14 sweet cherries and 6 sour cherries. If you pick a cherry at random, what is the probability it will be sweet?

Solution. Each of the cherries are equally likely to be selected since our selection is random and we can assume there is no way to distinguish one cherry from another. This means that the probability of selecting a sweet cherry will be equal to the number of sweet cherries in the bag dived by the total number of cherries in the bag. Since there are 14 sweet cherries and a total of 20 cherries in the bag, we have:

$$P(\text{sweet}) = \frac{14}{20} = \frac{7}{10} \text{ or } 70\%$$

There is a 70% chance of selecting a sweet cherry from the bag. □

4.2.6 Certain and Impossible Events

A probability is always a value between 0 and 1, or from 0% to 100%. If the probability of an event is 0 there are no outcomes that correspond with that event and we say it is ***impossible***. If the probability of an event is 1 then every outcome corresponds to that event and we say it is ***certain***.

Example 4.2.5

a. What is the probability of rolling an odd or even number on a six-sided die?

b. What is the probability of rolling an 8 on a six-sided die?

Solution.

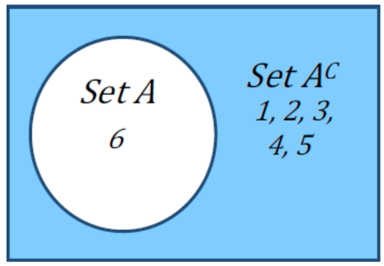
a. Since all the numbers are either even or odd, this event includes all of the outcomes in the sample space. This event is certain. $P(\text{odd or even}) = \frac{6}{6} = 1$

b. Since 8 is not one of the outcomes in the sample space, the event is impossible. $P(\text{roll an 8}) = \frac{0}{8} = 0$

□

4.2.7 Complementary Events

Just as we saw in the logic chapter, the ***complement*** of an event A means A does not happen. We can refer to the complement as ***not A*** or A^C. For example, consider the experiment of rolling a six-sided die and the simple event $\boldsymbol{A}$ = rolling a 6. The complement of event $\boldsymbol{A}$ is everything in the sample space that is not a 6: $A^C = 1, 2, 3, 4, 5$. Recall that we can illustrate the complement using a Venn diagram as shown below.



Notice that the outcomes from set A and the outcomes from set A^C will together equal the universal set, which is the sample space in probability. The probabilities must add up to 1 or 100%. Therefore, we can use subtraction to find the probability of a complement.

Complement of an Event.

$$P(A^C) = 1 - P(A)$$

Example 4.2.6 If you roll an eight-sided die, what's the probability you don't get a 6?

Solution. Not rolling a 6 is the complement of rolling a 6, which is easier to calculate. Since there are 8 possible numbers to roll, we have:

$$P(\text{not rolling a 6}) = 1 - P(\text{rolling a 6}) = 1 - \frac{1}{8} = \frac{7}{8} \text{ or } 0.875$$

There is an 87.5% chance of not rolling a 6. □

4.2.8 Experimental vs. Theoretical Probability

Now that we have calculated experimental and theoretical probabilities, we can compare them. When we flip a fair coin, we say there is a 50% chance of getting heads. This is a theoretical probability because there are two equally likely outcomes – heads and tails – so we expect to get heads half of the time. But if you flip a coin, say 100 times, will you get heads exactly 50 times? Maybe, but you are more likely to get some number around 50 times. The number of heads you actually observe out of the total number of times you flip the coin is the experimental probability.

Example 4.2.7 The table shows the numbers that came up after rolling a six-sided die 10 times. What is the experimental probability of rolling a 6? What is the theoretical probability of rolling a 6?

Roll	1	2	3	4	5	6	7	8	9	10
Outcome	3	1	4	6	6	6	1	3	5	1

Solution. To find the experimental probability of rolling a 6, it would be helpful to change this into a frequency table. We list all the possible outcomes and count how many times each occurred.

Outcome	***Frequency***
1	3
2	0
3	2
4	1
5	1
6	3

According to our frequency table, we see that a 6 was rolled three times, so the experimental probability of rolling a 6 is $P(\text{roll } 6) = \frac{3}{10}$ or 30%.

Theoretically, however, we would expect the number 6 to come up 1 out of 6 times since there are 6 equally likely outcomes. Thus, the theoretical probability of rolling a 6 is $P(\text{roll } 6) = \frac{1}{6}$ or approximately 16.7% □

4.2.9 The Law of Large Numbers

As we saw in the previous example, theoretical and experimental probabilities are not necessarily equal. However, experimental probability will eventually approach theoretical probability as we conduct more and more trials. This phenomenon is called the ***Law of Large Numbers***. This means if you flip a fair coin a small number of times, the experimental probability is likely to be different each time and could be very different from the theoretical probability. But if you flip a coin a large number of times, the experimental probability becomes very close to the theoretical probability of 50%. The Law of Large Numbers is extremely powerful in that it allows us to approximate the theoretical probability of complex events – like changes in beliefs and opinions, likelihood of natural disasters, climate change effects – through repeated sampling and simulation.

4.2.10 Probability of Compound Events

Now that we have the basics in place, let's look at some compound probability problems that we will be studying in this course.

4.2.11 "And" Probabilities

As we saw with truth tables, the event ***A and B*** refers to an event where both A and B occur. These events may occur at the same time or they could happen in a sequence such as ***A and then B***. How we calculate the theoretical probability of the event A *and* B (or A *and then* B) depends on whether the two events are independent or dependent.

4.2.12 Independent and Dependent Events

Two events A and B are ***independent*** if the probability of B occurring is the same whether or not A occurs. If the probability of B is affected by the occurrence of A, then we say that the events are ***dependent***.

Coin flips and die rolls are common examples of independent events – flipping heads does not change the probability of flipping heads the next time, nor does rolling a six change the probability that the next roll will be a six.

Another type of event is a selection event, such as randomly selecting or drawing items from a bag, etc. These are also independent if we ***draw with replacement***. By replacing the item, we reset the probability back to what it was before we made the selection. Since the probability of each subsequent selection is the same as the first selection, the events are independent.

If we draw ***without replacement***, however, like selecting multiple people for a committee, we change the total number of possible outcomes, thereby changing the probability of subsequent selections. Therefore, if we draw without replacement, the events will be ***dependent***.

Example 4.2.8 Determine whether the following events are independent or dependent.

a. Flipping a fair coin twice and getting heads both times.

b. Selecting a president and then a vice president at random from a pool of five equally qualified individuals.

c. The event that it will rain in Portland tomorrow and the event that it will rain in Beaverton tomorrow.

d. Wearing your lucky socks and getting an A on your exam.

Solution.

a. The probability of getting heads on the first flip is 0.5 or 50%. After flipping heads, the probability of getting heads on the second flip is still 0.5 or 50%. Since the probability of flipping heads on the second flip did not change because we flipped heads on the first flip, the events are ***independent***.

b. Since two different people will be put in the role of president and vice president, we are drawing without replacement and the events are therefore ***dependent***.

c. If it is raining in Portland it is more likely that it will rain in Beaverton, so the events are ***dependent***.

d. Although there may some sort of placebo effect at play in terms of confidence and persistence, the socks you wear do not have a direct effect on how well you do on your exam, so these events are ***independent***.

□

To calculate "and" probabilities we multiply, but we need to determine whether the events are independent or dependent. If they are independent, then we can multiply the individual probability of each event because one does not affect the other. If the events are dependent, then we need to multiply by the conditional probability based on what has previously happened. Here is a summary of this.

"And".

Probabilities If events A and B are independent, then $P(A \text{ and } B) = P(A) \cdot P(B)$

If events A and B are dependent, then $P(A \text{ and } B) = P(A) \cdot P(B \text{ given } A)$

The probability of *B given A* is called a ***conditional probability*** since it depends, or is conditional, on A occurring. We saw examples of conditional probability when we looked at contingency tables in the previous section.

Example 4.2.9 Suppose you have a bag containing 6 red Legos, 4 green Legos, and 3 black Legos. We will make exactly two lego selections from the bag. What is the probability of selecting:

a. two red Legos in a row if we put the first red Lego back in the bag?

b. two red Legos in a row if we don't put the first Lego back in the bag?

c. a red Lego and then a green Lego if we do not put the red Lego back in the bag?

Solution.

a. Since the outcomes are equally likely, the probability of selecting a red Lego is the number of red Legos divided by the total number of Legos, or $P(\text{red}) = \frac{6}{13}$.

If we replace the red Lego we selected (selections are independent), we go back to having 6 red Legos in the bag of 13 Legos total. Therefore,

$$\begin{aligned} P(\text{red and then red}) &= P(\text{red}) \cdot P(\text{red}) \\ &= \frac{6}{13} \cdot \frac{6}{13} \\ &\approx 0.213 \text{ or } 21.3\% \end{aligned}$$

b. If we do not replace the first red Lego (selections are dependent), then on our second draw there will only be 5 red Legos remaining, and 12 Legos in total. Therefore,

$$\begin{aligned} P(\text{red and then red}) &= P(\text{red}) \cdot P(\text{red given red taken out}) \\ &= \frac{6}{13} \cdot \frac{5}{12} \\ &\approx 0.192 \text{ or } 19.2\% \end{aligned}$$

c. The probability of selecting a red Lego on the first draw is the same as in parts a and b. Since we are not putting the red Lego back into the bag, we will have only 12 Lego left in total, of which 4 are green. Therefore,

$$\begin{aligned} P(\text{red and then green}) &= P(\text{red}) \cdot P(\text{green given red taken out}) \\ &= \frac{6}{13} \cdot \frac{4}{12} \\ &= \frac{6}{13} \cdot \frac{1}{3} \\ &\approx 0.154 \text{ or } 15.4\% \end{aligned}$$

□

Let's look at an example where we repeat an event many times.

Example 4.2.10 Suppose there is a 6% chance you will receive a citation if you ride the MAX train without a ticket. What is the probability that you get away without a single citation if you ride without purchasing a ticket for 20 days this month?

Solution. The first thing we want to recognize is that this question is essentially asking for the probability of no citation and no citation and no citation.... twenty times (one for each day you ride without buying a ticket). Since the outcomes are connected by an "and", we know we will be multiplying the probabilities. In this case the the outcomes are independent (you are not more or less likely to get a citation if you already received a citation). Therefore,

$$\begin{aligned} P(\text{no citation in 20 rides}) &= P(\text{no citation on a single ride})^{20} \\ &= (1 - 0.06)^{20} \\ &= (0.94)^{20} \\ &\approx 0.2901 \text{ or } 29.01\% \end{aligned}$$

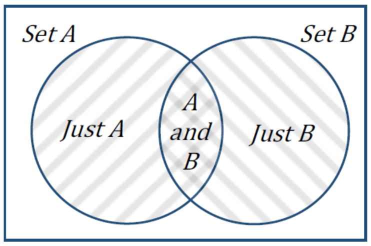
□

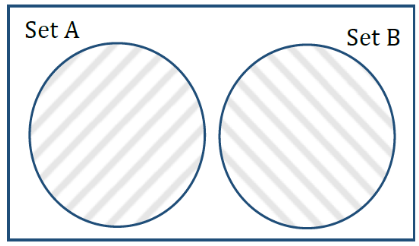
4.2.13 "Or" Probabilities

The event ***A or B*** refers to an event that includes the outcomes of A or B or *both*. We have seen the inclusive "or" both in terms of sets and logic, and in terms of contingency tables. The way we calculate the probability of A *or* B depends on whether the events have characteristics that are overlapping or disjoint.

4.2.14 Overlapping or Disjoint Sets

Recall that ***disjoint*** means the same thing as ***not overlapping***. Just like we saw in the logic and sets chapter, the set diagram on the left shows overlapping sets and the set diagram on the right shows disjoint sets.





To apply this to probability, we will look at an example of events that have overlapping characteristics, such as color and shape.

Example 4.2.11 A prize machine is filled with 10 yellow erasers, 6 green erasers, 4 red pencil sharpeners, 8 yellow pencil sharpeners, and 5 red bouncy balls. Each prize is inside a plastic sphere, and the spheres are well mixed in the prize machine. Each game will get you just one prize. Determine the probability of

a. getting a yellow prize.

b. getting a red or yellow prize.

c. getting a prize that is yellow or an eraser.

Solution.

a. Since yellow is a single event, we just need to know how many prizes there are in total, and how many of the prizes are yellow. The yellow prizes include the 10 yellow erasers and the 8 yellow pencil sharpeners.

$$P(\text{yellow}) = \frac{18}{33}$$

b. For a red or yellow prize, the set of red and the set of yellow do not overlap. They are disjoint sets, so we will add the probability of getting a red prize to the probability of getting a yellow prize.

$$\begin{aligned} P(\text{red or yellow}) &= P(\text{red}) + P(\text{yellow}) \\ &= \frac{9}{33} + \frac{18}{33} \\ &= \frac{27}{33} \end{aligned}$$

c. To find the probability of getting a prize that is yellow or an eraser, we need to be careful because these are overlapping sets. There are two ways to calculate this, and it is a lot like what we did with contingency tables. The first way is to add all the items separately, being careful not to double count.

$$
\begin{aligned}
P(\text{yellow or eraser}) &= P(\text{yellow eraser}) + P(\text{yellow pencil sharpener}) + P(\text{green eraser}) \\
&= \frac{10}{33} + \frac{8}{33} + \frac{6}{33} \\
&= \frac{24}{33}
\end{aligned}
$$

The second way is to count the total of yellow items and the total of erasers, but the yellow erasers are in both sets, or the overlap. We would be counting them twice and so we subtract their probability.

$$
\begin{aligned}
P(\text{yellow or eraser}) &= P(\text{yellow}) + P(\text{eraser}) + P(\text{yellow and eraser}) \\
&= \frac{18}{33} + \frac{16}{33} - \frac{10}{33} \\
&= \frac{24}{33}
\end{aligned}
$$

□

Here is a summary of how we found the "or" probabilities.

"Or" Probabilities.

If the sets are disjoint, $P(A \text{ or } B) = P(A) + P(B)$

If the sets are overlapping, $P(A \text{ or } B) = P(A) + P(B)\check{\ }P(A \text{ and } B)$

We could also use the overlapping formula as a general formula, because in the case of disjoint sets, there is no intersection and $P(A \text{ and } B) = 0$. Here is another example with overlapping events.

Example 4.2.12 What is the probability of rolling two fair dice and getting a pair or a sum of 6?

Solution. For complicated events it's a good idea to list all of the outcomes. Looking at the table of outcomes, we see that there are 6 outcomes that are pairs out of the 36 possible outcomes, and there are 5 outcomes that add to 6. We can also see that there is one outcome that is both a pair and a sum of 6, so the events are overlapping.

	1	2	3	4	5	6
1	$1+1=2$	$1+2=3$	$1+3=4$	$1+4=5$	$1+5=6$	$1+6=7$
2	$2+1=3$	$2+2=4$	$2+3=5$	$2+4=6$	$2+5=7$	$2+6=8$
3	$3+1=4$	$3+2=5$	$3+3=6$	$3+4=7$	$3+5=8$	$3+6=9$
4	$4+1=5$	$4+2=6$	$4+3=7$	$4+4=8$	$4+5=9$	$4+6=10$
5	$5+1=6$	$5+2=7$	$5+3=8$	$5+4=9$	$5+5=10$	$5+6=11$
6	$6+1=7$	$6+2=8$	$6+3=9$	$6+4=10$	$6+5=11$	$6+6=12$

As in the last example, there are two ways to do this.

If we add all of the shaded squares without double counting, we get:

$$
\begin{aligned}
P(\text{pair or sum of 6}) &= P(\text{pair}) + P(\text{sum of 6 that haven't been counted}) \\
&= \frac{6}{36} + \frac{4}{36} \\
&= \frac{10}{36}
\end{aligned}
$$

$$= \frac{5}{18}$$

To use the subtraction method, we need to add the probability of rolling a pair to the probability of rolling a sum of 6 and subtract the overlap. Thus we have:

$$\begin{aligned} P(\text{pair or sum of 6}) &= P(\text{pair}) + P(\text{sum of 6}) + P(\text{pair and a sum of 6}) \\ &= \frac{6}{36} + \frac{5}{36} - \frac{1}{36} \\ &= \frac{10}{36} \\ &= \frac{5}{36} \end{aligned}$$

Now that we have looked at empirical and theoretical probability, we will be able to use them for something very important in the next section – expected value. □

4.2.15 Exercises

1. A ball is drawn randomly from a jar containing 6 red marbles, 2 white marbles, and 5 yellow marbles. Find the probability of:

a. Drawing a white marble.

b. Drawing a red marble.

c. Drawing a green marble.

d. Drawing two yellow marbles if you draw with replacement.

e. Drawing first a red marble then a white marble if marbles are drawn without replacement.

2. Compute the probability of tossing a fair six-sided die and getting:

a. an even number.

b. a number less than 3.

3. Compute the probability of rolling a fair 12-sided die and getting:

a. a number other than 8.

b. a 2 or 7.

4. A fair six-sided die is rolled twice. What is the probability of getting:

a. a 6 on both rolls?

b. a 5 on the first roll and an even number on the second roll?

5. Suppose that 21% of people own dogs. If you pick two people at random, what is the probability that neither own a dog?

6. At some random moment, you look at your digital 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?

7. What is the probability of flipping a fair coin three times

a. and getting a head each time?

b. not getting a head at all?

8. What is the probability of rolling two fair six-sided dice

a. and getting a sum greater than or equal to 7?

b. getting an even sum or a sum greater than 7?

9. A box contains four black pieces of cloth, two striped pieces, and six dotted pieces. A piece is selected randomly and then placed back in the box. A second piece is selected randomly. What is the probability that:

a. both pieces are dotted?

b. the first piece is black, and the second piece is dotted?

c. one piece is black, and one piece is striped?

10. Compute the probability of rolling five fair six-sided dice (each side has equal probability of landing face up on each roll) and getting:

a. a 3 on all five dice.

b. at least one of the die shows a 3.

11. If you pick a card from a standard deck of 52 cards, what is the probability of getting

a. a 7?

b. a club?

c. a spade or a club?

d. a diamond or a 5?

12. A box of chocolates contains 7 dark chocolate pieces and 3 milk chocolate pieces (and no others). If you randomly pick 2 pieces and eat each chocolate after choosing it, what is the probability of choosing at least one dark chocolate? Write the probability in all three forms.

13. A bag contains 3 green marbles, 4 red marbles, and 5 blue marbles (and no others). If you randomly pull out three marbles all at once, what is the probability that you choose 3 blue marbles? Write the probability in all three forms.

14. A bag contains 3 green marbles, 4 red marbles, and 5 blue marbles (and no others). If you randomly pull out a marble and put the marble back 3 times, what is the probability that you pull out a blue marble all 3 times? Write the probability in all three forms.

15. A bag contains 3 green marbles, 4 red marbles, and 5 blue marbles (and no others). If you randomly pull out three marbles all at once, what is the probability that you choose at least 1 blue marble? Write the probability in all three forms.

4.3 Expected Value

Objectives: Section 4.3 Expected Value

Students will be able to:

- Make a probability model
- Calculate the expected value for a probability model
- Determine whether a game is fair

Expected value is one of the 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 may hope most people never learn about.

4.3.1 Expected Value

The ***expected value*** is the average gain or loss of an event if the procedure is repeated many times. To help get a better understanding of what expected value is and how it is used, consider the following scenario:

You are commissioned to design a game for a local carnival. Your proposed game will have players roll a six-sided die. If it comes up 6, they win $10. If not, they get to roll again. If they get a 6 on the second roll, then they win 3. If they do not get a 6 on the second roll, they lose. With the game design complete, you now need to decide how much the carnival game owner should charge players in order to make a profit over the long run.

To make a profit, the carnival needs to know how much they will pay in winnings, on average, over the long run and charge more than that. In other words, they must charge more than the expected value of the game.

One way to organize the outcomes and probabilities is with a probability model. A ***probability model*** or ***probability distribution*** is a table listing the possible outcomes and their corresponding probabilities. The outcomes will be the amounts a player can win, and we will calculate the probabilities using what we have learned about theoretical probability.

As we have seen with complements, probabilities in a probability distribution must add to 1, so that is important to check. Here is the probability model for the carnival game:

Outcome ($ won)	Rolling Event	Probability
$10	Roll a 6 on the first roll	$P(\text{roll a 6}) = \frac{1}{6}$
$3	Roll not a 6 on the first roll and a 6 on the second roll	$P(\text{roll a 6 then roll a 6}) = \frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36}$
$0	Roll not a 6 on the first roll and not a 6 on the second roll	$P(\text{roll a 6 then not a 6}) = \frac{5}{6} \cdot \frac{5}{6} = \frac{25}{36}$

Think of the expected value as a weighted average. We could take the average of $10, $3, and $0, but they are not all equally likely. It is much more likely to win $0 than to win $10. So, to find the average, we multiply each outcome by the chance it will happen and add the products together.

Expected Value.

Multiply each outcome by its probability and add up the products

In this case we have:

$$\text{Expected winnings} = \$10(\frac{1}{6}) + \$3(\frac{5}{36}) + \$0(\frac{25}{36}) = \$2.08$$

This tells us that over the long run, players can expect to win $2.08 per game. This also means that the carnival owner will be paying out an average of $2.08 per game! Since the carnival owner would rather not lose money by paying players over the long run we need to make sure to charge players enough to offset the average payout.

If the carnival owner charges exactly $2.08 to play, the game is considered a ***fair game*** since the expected winnings would be $0. In a fair game, the player isn't expected to win anything, nor is the owner expected to earn anything over the long run. However, if the carnival owner charges the player more than $2.08 to play, they will earn money over the long run.

Suppose you suggest charging $5 to play. We can determine the net winnings by subtracting the $5 the player has to pay from their expected winnings. This gives us:

$$\text{Net player winnings} = \$2.08 - \$5 = -\$2.92$$

This means that over the long run, players can expect to lose an average of $2.92 each game they play, and the carnival owner can expect to earn an average of $2.92 per game over the long run. Here's another example.

Example 4.3.1 Pick4 is a game by the Oregon Lottery that costs $1 to play. In this game you pick 4 numbers in a specific pattern. If you get the exact sequence, you can in theory earn a lot of money. Suppose that the payouts are as follows. Determine the player's expected net winnings.

Prize ($)	***Probability***
$250	$^1/_{417}$
$500	$^1/_{1833}$
$1000	$^1/_{1667}$
$1500	$^1/_{2500}$

Solution. This table is not quite a complete probability distribution since it is missing one important outcome: when the player loses. In that case the prize is $0. We need to add a line for this. The prize for this missing outcome is $0, and since losing is the complement to winning *something*, the probability will be:

$$\begin{aligned} P(\text{Win } \$0) &= 1 - (\frac{1}{417} + \frac{1}{1833} + \frac{1}{1667} + \frac{1}{2500}) \\ &= 1 - 0.0039 \\ &= 0.9961 \end{aligned}$$

Adding this information to the table gives a complete probability distribution. Now we can see that players are going to lose more than 99% of the time, so the expected value will be heavily weighted toward winning $0.

Prize ($)	***Probability***
$250	$^1/_{417}$
$500	$^1/_{1833}$
$1000	$^1/_{1667}$
$1500	$^1/_{2500}$
$0	0.9961

$$\text{Expected winnings} = \$250(\frac{1}{417}) + \$500(\frac{1}{1833}) + \$1000(\frac{1}{16773}) + \$1500(\frac{1}{2500}) + \$0(0.9961)$$

$$= \$2.07$$

Therefore, the player's expected winnings are \$2.07, on average, over the long run. To find the expected ***net*** winnings, we subtract the cost to play. Since it costs \$1 to play,

$$\text{Net Expected Winnings} = \$2.07 - \$1.00 = \$1.07.$$

Assuming the given payouts are correct, this would be one game you would want to play for investment purposes since you can expect to earn \$1.07 per game, on average, over the long run. Play a million times, and you just might become a millionaire! □

In general, if the expected value of a game is negative, it is not a good idea to play, since in the long run 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.

Not surprisingly, the expected value for casino games is always negative for the player, and therefore positive for the casino. It must be positive for the casino, or they would go out of business! Players just need to keep in mind that when they play a game repeatedly, they should expect to lose money. That is fine so long as you enjoy playing the game and think it is worth the cost, but it would be wrong to expect to come out ahead.

Expected value is not only used to determine the average amount won and lost at casinos and carnivals, it also has applications in business and insurance, just to name a few. Let's look at a couple of those applications.

Example 4.3.2 For 3 months, a coffee shop tracked their morning sales of coffee, between 6am and 10am. The following results were recorded:

Number of cups sold	145	150	155	160	170
Probability	0.15	0.22	0.37	0.19	0.07

How many cups of coffee should they expect to sell each morning?

Solution. In this case the table tells us that 15% of the time they sell 145 cups of coffee between 6am and 10am, 22% of the time they sell 150 cups, 37% of the time they sell 155 cups, 19% of the time they sell 160 cups, and 7% of the time they sell 170 cups. Since the highest probability is associated with 155 cups, the expected value should lie somewhat close to this.

To find the expected number of coffees sold, we multiply each number of cups of coffee by its respective probability and then add the products.

$$\begin{aligned}\text{Expected number of coffees sold} &= 145(0.15) + 150(0.22) + 155(0.37) + 160(0.19) + 170(0.07)\\ &= 154.4 \text{ cups of coffee}\end{aligned}$$

This means that over the long run, the coffee shop can expect, on average, to sell around 154 cups of coffee each morning. This is an important tool for businesses since it helps inform them how much stock they should keep on hand. □

Example 4.3.3 On average, a 40-year-old man in the US has a 0.242% chance of dying in the next year[1]. An insurance company charges \$275 annually for a life insurance policy that pays a \$100,000 death benefit. What is the expected value for the insurance company on this policy?

Solution. The first thing we want to do is organize the probabilities and outcomes in a probability distribution table. There are two outcomes – either the policy holder dies, and the insurance company pays the benefit, or the policy holder does not die, and they do not pay anything.

The probability of paying the death benefit is equal to the chance of the person dying in the next year, and the probability of paying nothing is equal to the complement of the chance of dying in the next year.

Insurance payout	***Probability***
\$100000	0.00242
\$0	1-0.00242=0.99758

Then we can calculate:

$$\text{Expected Payout} = \$100000(0.00242) + \$0(0.99758) = \$242$$

So, the expected payout for the insurance company is \$242, but they are charging \$275 for the policy. Their net revenue would be

$$\text{Net Value to Insurance Company} = \$275 - \$242 = \$33$$

The insurance company is making, on average, \$33 per policy per year. It shouldn't be too surprising because there are overhead costs and the insurance company can only afford to offer policies if they, on average, make money on them. But how much money should they make? As a consumer it is important to know about expected value. □

4.3.2 Exercises

1. 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.

a. Make a probability model for this game.

b. What is your expected value if you play this game?

2. A bag contains 5 red marbles, 8 blue marbles, and 15 green marbles. Someone offers to play this game: You randomly select one marble from the bag. If it is blue, you win \$3. If it is red, you win \$2. If it is green, you lose \$1.

a. Make a probability model for this game.

b. What is your expected value if you play this game?

3. A friend devises a game that is played by rolling a single six-sided fair (each side has equal probability of landing face up, once rolled) 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.

a. Make a probability model for this game.

b. Compute the expected value for this game.

c. Should you play this game?

4. A friend devises a game that is played by rolling a single six-sided fair (each side has equal probability of landing face up, once rolled) die once. If you roll a 1, he pays you \$5; if you roll a 2, he pays you nothing; if you roll a number more than 2, you pay him \$2.

a. Make a probability model for this game.

b. Compute the expected value for this game.

[1] According to the estimator at http://www.numericalexample.com/index.php?view=article;&ampid=91

c. Should you play this game?

5. A company wants to offer a 2-year extended warranty in case their product fails after the original warranty period but within 2 years of the purchase. They estimate that 0.7% of their products will fail during that time, and it will cost them \$350 to replace a failed product. If they charge \$48 for the extended warranty, what is the company's expected profit or loss on each warranty sold?

6. A company wants to offer a 3-year extended warranty in case their product fails after the original warranty period but within 3 years of the purchase. They estimate that 0.5% of their products will fail during that time, and it will cost them \$400 to replace a failed product. If they charge \$60 for the extended warranty, what is the company's expected profit or loss on each warranty sold?

7. 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?

8. An insurance company estimates the probability of a flood in the next year to be 0.0002. The average damage done by a flood is estimated to be \$70,000. If the company offers flood insurance for \$3000, what is their expected value of the policy?

9. 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.

10. You purchase a raffle ticket to help out a charity. The raffle ticket costs \$10. The charity is selling 1000 tickets. One of them will be drawn and the person holding the ticket will be given a prize worth \$3000. Compute the expected value for this raffle.

11. At the local fair there is a game in which folks are betting where a chicken will poop on a 5 by 5-foot grid. (There are 25, 1 by 1 squares to choose from) You can buy a 1 by 1-foot square for \$10 and if the chicken poops on your square you win \$100. Find the expected value for this game.

12. At the local fair there is a game in which folks are betting where a chicken will poop on a 4 by 4-foot grid. (There are 16, 1 by 1 squares to choose from) You can buy a 1 by 1-foot square for \$15 and if the chicken poops on your square you win \$125. Find the expected value for this game.

13. Create a problem using the concept of expected value. Possible topics include insurance policies, financial decisions, gambling and lotteries. Determine the expected value of the situation you created.

4.4 Chapter 4 Review

Review Exercises

1. A professor gave a test to students in a morning class and the same test to the afternoon class. The grades are summarized below.

	A	B	C	Total
Morning Class	14	11	7	32
Afternoon Class	11	13	4	28
Total	25	24	11	60

If one student was chosen at random, find each probability:

a. P(in the afternoon class)

b. P(earned an A)

c. P(earned a B and was in the afternoon class)

d. P(earned a C given the student was in the morning class)

e. P(is in the morning class given that the student earned a B)

2. A professor gave a test to students in a science class and in a math class during the same week. The grades are summarized below.

	A	B	C	Total
Science Class	7	18	13	38
Math Class	10	8	9	27
Total	17	26	22	65

If one student was chosen at random, find each probability:

a. P(in the math class)

b. P(earned a B)

c. P(earned an A and was in the math class)

d. P(earned a B given the student was in the science class)

e. P(is in the math class given that the student earned a B)

3. Four Cable Channels (2, 6, 8, 12) have Drama series, Sitcoms, Game Shows, and News. The number of each type of show is listed in the table below. Complete the table and use it to answer the questions.

Type	Channel 2	Channel 6	Channel 8	Channel 12	Total
Drama	5	2	4		15
Sitcom	6		7	3	25
Game Show		4	3	4	15
News	3	2		3	10
Total					65

a. P(Sitcom or Game Show)

b. P(Drama and Channel 8)

c. P(Channel 8 or Channel 2)

d. P(Drama given that it is on Channel 6)

e. Given that the show is a sitcom, find the probability it is on channel 12.

f. Find the probability that a show is a game show, given that the show is on channel 2.

4. Four Cable Channels (3, 5, 7, 13) have Reality shows, Crime dramas, Cooking Shows, and Community Programming. The number of each type of show is listed in the table below. Complete the table and use it to answer the questions.

Type	Channel 3	Channel 5	Channel 7	Channel 13	Total
Reality	6		8	6	
Crime	2	1		2	
Cooking	5	5	9		11
Community		8	4		
Total	15	21	25		83

a. P(Reality or Crime Show)

b. P(Cooking and Channel 3)

c. P(Channel 3 or Channel 13)

d. P(Community Program given that it is on Channel 5)

e. Given that the show is a crime show, find the probability it is on channel 3.

f. Find the probability that a show is a cooking show, given that it is on channel 7.

5. A ball is drawn randomly from a jar containing 12 red marbles, 8 white marbles, and 5 yellow marbles. Find the probability of:

a. Drawing a red marble.

b. Not drawing a white marble.

c. Drawing a yellow or red marble.

d. Drawing a blue marble.

e. Drawing two red marbles if you draw with replacement.

f. Drawing first a red marble then a yellow marble if marbles are drawn without replacement.

6. A ball is drawn randomly from a jar containing 18 black marbles, 4 purple marbles, and 9 blue marbles. Find the probability of:

a. Drawing a black marble.

b. Not drawing a purple marble.

c. Drawing a blue or purple marble.

d. Drawing a yellow marble.

e. Drawing two black marbles if you draw with replacement.

f. Drawing first a blue marble then a black marble if marbles are drawn without replacement.

7. What is the probability of flipping a coin four times

a. and getting a head each time?

b. not getting a head at all?

8. What is the probability of flipping a coin 7 times

a. and getting all tails?

b. getting all heads?

9. According to a survey by Pew Research in 2020, 68% of U.S. adults say the federal government is doing too little to protect water quality. (+/- 1.6%)[1] If you pick two adults at random, what is the probability that

a. Both of them think the government is doing too little to protect water quality.

b. Neither of them thinks the government is doing too little to protect water quality.

10. According to a national AP-NORC Survey, 95% of U.S. adults think changes are needed in the criminal justice system (+/- 3.7%).[2] If you pick 3 people at random, what is the probability that

a. All of them support criminal justice reform.

b. None of them support criminal justice reform

11. A bag contains 2 black marbles, 4 orange marbles, and 20 yellow marbles. Someone offers to play this game: You randomly select one marble from the bag. If it is black, you win \$3. If it is orange, you win \$2. If it is yellow, you lose \$1.

a. Make a probability model for this game.

b. What is your expected value if you play this game?

c. Should you play this game?

12. A friend devises a game that is played by rolling a single six-sided die once. If you roll a 6, he pays you \$10; if you roll a 5, he pays you nothing; if you roll a number less than 5, you pay him \$1.

a. Make a probability model for this game.

b. Compute the expected value for this game.

c. Should you play this game?

13. A company wants to offer a 2-year extended warranty in case their product fails after the original warranty period but within 2 years of the purchase. They estimate that 1.5% of their products will fail during that time, and it will cost them \$450 to replace a failed product. If they charge \$55 for the extended warranty, what is the company's expected profit or loss on each warranty sold?

14. You purchase a raffle ticket to help out a charity. The raffle ticket costs \$10. The charity is selling 2000 tickets. One of them will be drawn and the person holding the ticket will be given a prize worth \$8000. Compute the expected value for this raffle.

[1] https://www.pewresearch.org/fact-tank/2020/04/21/how-americans-see-climate-change-and-the-environment-in-7-charts/
[2] http://www.apnorc.org/projects/Pages/Widespread-Desire-for-Policing-and-Criminal-Justice-Reform.aspx

Chapter 5

Democracy

Dr. Graciela Chichilnisky and Social Choice.

Dr. Graciela Chichilnisky[1] is a professor of Economics and Mathematical Statistics at Columbia University. She holds Ph.D. degrees in Mathematics and Economics from MIT and the University of California Berkeley.

Dr. Chichilnisky applied algebraic topology to voting theory and the theory of social choice. She accounted for not only the order of voting preference but also the intensity of preference among the alternatives (Chichilnisky, 1980). Due to her innovative work in this area, continuous social choice has become an international subdiscipline.

More recently, Dr. Chichilnisky combined her research in math and economics to focus on climate change. Dr. Chichilnisky acted as the lead author on the Intergovernmental Panel on Climate Change, which received the 2007 Nobel Prize for its work in deciding world policy with respect to climate change, and she worked extensively on the Kyoto Protocol, creating and designing the carbon market that became international law in 2005 (Siegle, 2010).

As a world-renowned economist, she is the creator of the formal theory of Sustainable Development (Siegle, 2010). Her pioneering work uses innovative market mechanisms to create Green Capitalism.

Dr. Chichilnisky is the CEO and Co-founder of Global Thermostat[2], a company in which she co-invented a "Carbon Negative Technology" that captures CO2 from the air and transforms it into profitable assets such as biofuels, food, beverages, polymers, and valuable building materials (Chichilnisky, 2018).

Dr. Chichilnisky taught previously at Harvard, Essex and Stanford Universities. She was elected one of the Ten Most Influential Latinos in the U.S. She is a frequent political and economic speaker on CNN, ABC, BBC TV News, and Bloomberg News, as well as a frequent keynote speaker at leading international conferences and universities. Her two most recent books are "Reversing Climate Change: How Carbon Removals Can Resolve Climate Change and Fix the Economy" and "The Economics of Climate Change". Dr. Chichilnisky is also the author of "Saving Kyoto", which won the American Library Association's 2010 Outstanding Academic Title of the Year and the American Geographical Society's Book of the Month Award in October 2009.

[1] Photo: ©Michael Timmons. All rights reserved, used with permission.
[2] https://globalthermostat.com/

5.1 Apportionment

Objectives: Section 5.1

Students will be able to:

- Explain the historical context of the formation of the U.S Government, including colonization, slavery and the Three-fifths Compromise, and how they relate to systemic inequality today
- Apportion representatives using Hamilton's method
- Apportion representatives using Jefferson's method
- Apportion representatives using Webster's method
- Apportion representatives using the Hill-Huntington Method

5.1.1 Historical Context for This Chapter

In this chapter we are going to study some of the math used in the United States government today. In doing this, it is very important to acknowledge the Native American genocide and race-based chattel slavery which are often left out of math textbooks.

We are including this history to show how racist policies, laws and practices like segregation, racial profiling, police brutality, mass incarceration, redlining and voter suppression continue to this day to limit the freedoms, rights and economic prospects of African Americans. With the COVID-19 pandemic and the death of George Floyd by a police officer, there is greater awareness of systemic racism than ever before.

5.1.2 Formation of the United States

Scholars estimate that people have lived in the Americas for over 20,000 years (Worrall, 2018). By the time Columbus arrived, more than 50 million people were living in the Americas and around 10 million Indegenous people were living in the area that would become the United States (History.com Editors, 2018).

The first colonists arrived in America in 1607 and African people who had been violently separated from their families and homeland were brought to Jamestown in 1619 as slaves. Slave labor was used on tobacco and cotton plantations, for domestic work and even as part of the labor for building the White House. The system of slavery was an integral part of the United States economic and political system (How Slavery Helped Build a World Economy, 2003). At the same time, white settlements kept encroaching on the homeland of Native Americans. Many treaties were signed and then violated. For example, in 1785, the ***Treaty of Hopewell*** was signed with the Cherokees in Georgia, which set a boundary on white settlement, but there were already colonists past that border.

The U.S. government was officially formed in 1787. During the ***Constitutional Convention***, the framers were tasked with making decisions on the three branches of government and how people (or those considered to be people) would be represented. At that time only white men with property were allowed to vote. Women were also considered to be property of their husband at the time.

In creating the Legislative Branch of government, it was decided that Congress would be made up of the Senate and the House of Representatives. Representation for the ***Senate*** would be ***equal*** with two senators per state. To account for the different sizes of the states, the representation in the ***House of Representatives*** would be ***proportional*** to the state's population, but there was a controversy over how to count the population.

During the Constitutional Convention, the northern and southern states were at odds on whom to count. The South wanted more representation in Congress which would also increase their power, so they argued to count slaves as part of their population. The southern states did not want to count slaves for taxation, however, because that would result in them paying more taxes.

Slaves were considered property, not humans, and they were not citizens. Therefore, they did not have the right to vote or participate in government. The anti-slavery North wanted to count only free people, which included free Blacks. This led to the ***Three-Fifths Compromise*** that determined three out of every five slaves would be counted toward a state's population and taxation (Clayton, 2015).

It is also important to note that there are about 4 million U.S. Citizens who live in the territories of Guam, the Virgin Islands, the Northern Mariana Islands and Puerto Rico who do not have representatives or senators. They pay federal taxes like Social Security and Medicare but not Federal Income tax.

5.1.3 What is Apportionment?

The number of representatives each state gets is based on its population, so once the framers decided how to count the population, they had to figure out how to divide up the representatives. This math problem is called apportionment.

Apportionment is the problem of dividing up a fixed number of things among groups of different sizes. In the United States, there is a certain number of representatives as stated in the constitution, currently 435, and they need to be divided fairly among the 50 states. Since the states are different sizes, and we cannot use fractions of people, this is not an easy task. In addition, the population in each state may change over time. Every 10 years after the census is taken, the representatives are reapportioned.

The apportionment problem comes up in a variety of non-political areas too. Here are the rules for apportionment in general.

Apportionment rules.

1. The things being divided up can exist only in whole numbers.
2. We must use all of the things being divided up, and we cannot use any more.
3. Each group must get at least one of the things being divided up.
4. The number of things assigned to each group should be at least approximately proportional to the population of the group. (Exact proportionality isn't possible because of the whole number requirement, but we should try to be close.)

In terms of the apportionment of the United States House of Representatives, these rules imply:

1. We can only have whole representatives (a state can't have 3.4 representatives).
2. We can only use the 435 representatives available.
3. Every state gets at least one representative.
4. The number of representatives each state gets should be approximately proportional to the state population. This way, the number of constituents each representative has should be approximately equal.

We will look at four ways of solving the apportionment problem, developed by different people: Hamilton, Jefferson, Webster, and the Huntington-Hill method that is used today. We will continue to look at U.S. history throughout this section including Hamilton, Jefferson and Webster's relationship to slavery.

5.1.4 Hamilton's Method

Alexander Hamilton (1755 - 1804) was raised in St. Croix in the U.S. Virgin Islands by a poor family. He was motivated by his low socioeconomic status to work himself into a higher social standing. He eventually came to the colonies and worked himself into circles of wealth and influence. While Hamilton wasn't pro-slavery and considered himself an abolitionist, when choosing between his societal status and moral obligation, he chose the former.

He wed Elizabeth Schuyler, who was from a prominent family who owned slaves. He was involved with the transactions of slaves for his in-laws which further muddled his anti-slavery stance. Furthermore, Hamilton also traded and sold slaves as part of his duties for the Continental Army.

Alexander Hamilton proposed the method that now bears his name. His method was approved by Congress in 1791 but was vetoed by President Washington. It was later adopted in 1852 and used through 1910. He begins by determining, to several decimal places, how many people each representative should represent (the divisor).

Hamilton's Method.

1. Determine how many people each representative should represent. Do this by dividing the total population of all the states by the total number of representatives. This answer is called the ***divisor***.

2. Divide each state's population by the divisor to determine how many representatives it should have. Record this answer to several decimal places. This answer is called the ***quota***.

   Since we can only allocate whole representatives, Hamilton resolved the whole number problem as follows:

3. Cut off all the decimal parts of all the quotas (but don't forget what the decimals were). This is the ***initial*** apportionment and will always be less than or equal to the total number of representatives. Add up the whole numbers.

4. If the total from Step 3 is less than the total number of representatives, assign the remaining representatives, one each, to the states whose decimal parts of the quota were largest, until the desired total is reached.

   Make sure that each state ends up with at least one representative.

Let's see how this works in an example.

Example 5.1.1 A new state is named after George Floyd, killed by the police in Minneapolis, Minnesota in 2020. Floyd has three counties: King, Garner and Taylor. They are named after African Americans who were also victims of police brutality. Rodney King was brutally beaten by four white police officers and Eric Garner and Breonna Taylor were killed by the police. The Floyd State House of Representatives has 41 members. If the legislature wants to divide this representation along county lines (which is *not* required, but let's pretend they do), let's use Hamilton's method to apportion them. The populations of the counties are as follows:

County	Population
King	162,310
Garner	538,479
Taylor	197,145
Total	897,934

Step 1: First, we divide the total population by the number of representatives to find the divisor: $897934/41 = 21900.82927$.

Step 2: Now we determine each county's quota by dividing the county's population by the divisor: For example, for King, we take $162,310/21,900.82927$ which equals 7.4111.

County	Population	Quota
King	162,310	7.4111
Garner	538,479	24.5872
Taylor	197,945	9.0017
Total	897,934	

Step 3: Removing the decimal parts of the quotas gives our initial apportionment and we add those numbers up.

County	Population	Quota	Initial
King	162,310	7.4111	7
Garner	538,479	24.5872	24
Taylor	197,945	9.0017	9
Total	897,934		40

Step 4: We need 41 representatives, and right now we only have 40. The remaining one goes to the county with the largest decimal part, which is Garner:

County	Population	Quota	Initial	Final
King	162,310	7.4111	7	7
Garner	538,479	24.*5872*	24+*1*	25
Taylor	197,945	9.0017	9	9
Total	897,934		40	41

Our final apportionment is King: 7, Garner: 25, Taylor: 9, for a total of 41 representatives. □

Example 5.1.2 We will use Hamilton's method again, to apportion 75 seats for a new state of Lewis, which has five counties. Lewis and its counties are named for civil rights leaders John Lewis, Rosa Parks, Dr. Martin Luther King Jr., Ella Baker, Daisy Bates and Roy Wilkins.

Solution. Step 1: The divisor is $1,052,567/75 = 14,034.22667$.

Step 2: Determine each county's quota by dividing its population by the divisor:

County	Population	Quota
Parks	49,875	3.5538
King	166,158	11.8395
Baker	82,888	5.9061
Bates	626,667	44.6528
Wilkins	126,979	9.0478
Total	1,052,567	

Step 3: Remove the decimal part of each quota and add up the initial apportionment:

County	Population	Quota	Initial
Parks	49,875	3.5538	3
King	166,158	11.8395	11
Baker	82,888	5.9061	5
Bates	626,667	44.6528	44
Wilkins	126,979	9.0478	9
Total	1,052,567		72

Step 4: We need 75 representatives and we only have 72, so we assign the remaining three, one each, to the three counties with the largest decimal parts, which are Baker, King, and Bates in that order:

County	Population	Quota	Initial	Final
Parks	49,875	3.5538	3	3
King	166,158	11.*8395*	11+*1*	12
Baker	82,888	5.*9061*	5+*1*	6
Bates	626,667	44.*6528*	44+*1*	45
Wilkins	126,979	9.0478	9	9
Total	1,052,567		72	75

Our final apportionment is Parks: 3, King: 12, Baker: 6, Bates: 45, and Wilkins: 9 for a total of 75 representatives.

Note that even though Parks County's decimal part is greater than .5, it isn't big enough to get an additional representative, because three other counties have greater decimal parts. □

Hamilton's method obeys something called the Quota Rule. The Quota Rule isn't a law, but an idea that some people think is a good one.

Quota Rule.

The ***Quota Rule*** says that the final number of representatives a state gets should be within one of that state's quota. Since we're dealing with whole numbers for our final answers, that means that each state should either go up to the next whole number above its quota, or down to the next whole number below its quota.

5.1.5 Problems with Hamilton's Method

After using Hamilton's method for many years, three paradoxes happened, on separate occasions, where unfair things happened in new apportionments. This led to other methods being needed.

The Alabama Paradox is named for an incident that happened during the apportionment that took place after the 1880 census. (A similar incident happened ten years earlier involving the state of Rhode Island, but the paradox is named after Alabama.) The post-1880 apportionment had been completed, using Hamilton's method and the new population numbers from the census. Then it was decided that because of the country's growing population, the House of Representatives should be made larger. That meant that the apportionment would need to be done again, still using Hamilton's method and the same 1880 census numbers, but with more representatives. The assumption was that some states would gain another representative and others would stay with the same number they already had (since there weren't enough new representatives being added to give one more to every state). The paradox is that Alabama ended up *losing* a representative in the process, even though no populations were changed, and the total number of representatives increased.

The New States Paradox happened when Oklahoma became a state in 1907. Oklahoma had enough population to qualify for five representatives in Congress. Those five representatives would need to come from somewhere, though, so five states, presumably, would lose one representative each. That happened, but another thing also happened: Maine gained a representative (from New York).

The Population Paradox happened between the apportionments after the census of 1900 and of 1910. In those ten years, Virginia's population grew at an average annual rate of 1.07%, while Maine's grew at an average annual rate of 0.67%. Virginia started with more people, grew at a faster rate, grew by more people, and ended up with more people than Maine. By itself, that doesn't mean that Virginia should gain representatives or Maine shouldn't, because there are lots of other states involved. But Virginia ended up losing a representative *to Maine*.

5.1.6 Jefferson's Method

"All men are created equal," are words penned by our third president and Founding Father, Thomas Jefferson (1743 - 1826), in the preamble to the Constitution. However, over the course of his life, Jefferson owned around 600 slaves. Among these hundreds of slaves, Jefferson fathered at least six children with one of his slaves, Sally Hennings.

Over his lifetime and his writings, Jefferson wrestled with his conscience regarding slavery, which can be seen in documents such as drafts of the Constitution before its final version. In the end, his personal and monetary gain surpassed that of his concern for the enslaved.

Thomas Jefferson proposed a different method for apportionment. After Washington vetoed Hamilton's method, Jefferson's method was adopted, and used in Congress from 1791 through 1842. Jefferson, of course, had political reasons for wanting his method to be used rather than Hamilton's. Primarily, his method favors larger states, and his own home state of Virginia was the largest at the time. He would also argue that it's the ratio of people to representatives that is the critical thing, and apportionment methods should be based on that.

The first three steps of Jefferson's method are the same as Hamilton's. He found the same divisor and the same quota and cut off the decimal parts in the same way, giving the same initial apportionment that is less than the required total.

What changes is how Jefferson assigned the remaining representatives. He said that since we ended up with an answer that is too small, our divisor must have been too big. He changed the divisor by making it smaller and looked at the new total. He would raise or lower the divisor until he found one that produced the required total. This is a trial and error process that takes some patience.

Jefferson's Method.

The first three steps are the same as Hamilton's

1. Determine how many people each representative should represent. Do this by dividing the total population of all the states by the total number of representatives. This answer is called the ***divisor***.
2. Divide each state's population by the divisor to determine how many representatives it should have. Record this answer to several decimal places. This answer is called the ***quota***.
3. Cut off all the decimal parts of all the quotas (but don't forget what the decimals were). This is the ***initial*** apportionment and will always be less than or equal to the total number of representatives. Add up the whole numbers.
4. If the total from Step 3 was less than the total number of representatives, reduce the divisor and recalculate the quota and allocation. If you lower it too far, increase it. Continue doing this until the total in Step 3 is equal to the total number of representatives. The divisor you end up using is called the ***modified divisor***. This is a trial-and-error process.

Example 5.1.3 We'll return to the state of Floyd and apply Jefferson's method. We begin, as we did with Hamilton's method, by finding the quotas with the original divisor, 21,900.82927.

Steps 1-3 are the same as Hamilton's Method. Since we will be using different divisors, we will write the divisor above the population column to be clear what we are dividing each number in that column by.

	÷21,900.82927		
County	Population	Quota	Initial
King	162,310	7.4111	7
Garner	538,479	24.5872	24
Taylor	197,145	9.0017	9
Total	897,934		40

We need 41 representatives, and this divisor gives only 40. We must reduce the divisor until we get 41 representatives. Let's try 21,500 as the divisor

Step 4:

	÷21,500		
County	Population	Quota	Initial
King	162,310	7.5493	7
Garner	538479	25.0455	25
Taylor	197,145	9.1695	9
Total	897,934		41

That worked and our final apportionment is King: 7, Garner: 25, and Taylor: 9.

Notice that with the new, lower divisor, the quota for Garner County (the largest county in the state) increased by much more than those of King County or Taylor County.

In this example, we got lucky and found the modified divisor on the first try. If we still had 40, we would reduce the divisor more. If we had more than 41, we would need to raise it. We will show how to do the trial-and-error part in the next example. □

Example 5.1.4 We'll apply Jefferson's method for Lewis. The original divisor of 14,034.22667 gave these results:

	÷14,034.22667		
County	Population	Quota	Initial
Parks	49,875	3.5538	3
King	166,158	11.8395	11
Baker	82,888	5.9061	5
Bates	626,667	44.6528	44
Wilkins	126,979	9.0478	9
Total	1,052,567		72

We need 75 representatives and we only have 72, so we need to use a smaller divisor. Let's try lowering it to 13,500:

	÷13,500		
County	Population	Quota	Initial
Parks	49,875	3.6944	3
King	166,158	12.3080	12
Baker	82,888	6.1399	6
Bates	626,667	46.4198	46
Wilkins	126,979	9.4059	9
Total	1,052,567		76

We got a total of 76 representatives which is too many, so we lowered it too far. We need a divisor that's greater than 13,500 but less than 14,034.22667. Let's try 13,700:

	÷13,700		
County	Population	Quota	Initial
Parks	49,875	3.6405	3
King	166,158	12.1283	12
Baker	82,888	6.0502	6
Bates	626,667	45.7421	45
Wilkins	126,979	9.2685	9
Total	1,052,567		75

Using a modified divisor of 13,700 gives us exactly 75 representatives. Note there is usually more than one modified divisor that will work.

This can take a lot of writing, so in practice, we can write this all out in one table. With each try we are dividing the population by the new divisor to get the new quotas.

	÷14,034.22667			*÷13,500*		*÷13,700*	
County	Population	Quota	Initial	2nd Quota	2nd Try	3rd Quota	Final
Parks	49,875	3.5538	3	3.6944	3	3.6405	3
King	166,158	11.8395	11	12.3080	12	12.1283	12
Baker	82,888	5.9061	5	6.1399	6	6.0502	6
Bates	626,667	44.6528	44	46.4198	46	45.7421	45
Wilkins	126,979	9.0478	9	9.4059	9	9.2685	9
Total	1,052,567		72		76		75

Notice, in comparison to Hamilton's method, that although the results were the same, they came about in a different way, and the outcome was almost different. Bates County (the largest) almost went up to 46 representatives before King (which is much smaller) got to 12. Although that didn't happen here, it can. Divisor-adjusting methods like Jefferson's are not guaranteed to follow the quota rule. □

5.1.7 Webster's Method

Daniel Webster (1782-1852) was a lawyer, congressman, Senator of Massachusetts and also served as the U.S. Secretary of State. He was from the North and did not own slaves. He was an opponent of slavery extension and he spoke against annexing Texas and against going to war with Mexico. He argued, however, that no law was needed to prevent the further extension of slavery in new states and he supported the Compromise of 1850, which disappointed his abolitionist supporters (History.com Editors, 2018b).

Webster proposed a method similar to Jefferson's in 1832. It was adopted by Congress in 1842 but replaced by Hamilton's method in 1852. It was then adopted again in 1911. The difference is that Webster rounded the quotas to the nearest whole number rather than dropping the decimal parts. If that didn't produce the exact number of representatives, he adjusted the divisor like in Jefferson's method. (In Jefferson's case, the first adjustment will always be to make the divisor smaller. That is not always the case with Webster's method because some numbers may be rounded up.)

Webster's Method.

Steps 1-2 are the same as Hamilton and Jefferson

1. Determine how many people each representative should represent. Do this by dividing the total population of all the states by the total number of representatives. This answer is called the ***divisor***.

2. Divide each state's population by the divisor to determine how many representatives it should have. Record this answer to several decimal places. This answer is called the ***quota***.

3. Round all the quotas to the nearest whole number (but don't forget what the decimals were). This is the ***initial*** apportionment. Add up the whole numbers.

4. If the total from Step 3 is less than the total number of representatives, reduce the divisor and recalculate the quota and allocation. If the total from step 3 is larger than the total number of representatives, increase the divisor and recalculate the quota and allocation. Continue doing this until the total in Step 3 is equal to the total number of representatives. The divisor we end up using is called the ***modified divisor***. This is a trial-and-error process.

Let's see how Webster's method works in this example:

Example 5.1.5 We will look at Floyd again, with an initial divisor of 21,900.82927 and 41 representatives:

	÷21,900.82927		
County	Population	Quota	Initial
King	162,310	7.4111	7
Garner	538,479	*24.5872*	25
Taylor	197,145	9.0017	9
Total	897,934		41

This time Garner is the only county with a decimal of 0.5 or higher, so it gets rounded up. This gives the required total, so we're done. □

Example 5.1.6 Let's look at Lewis again, with an initial divisor of 14,034.22667 and 75 representatives:

	÷14,034.22667		
County	Population	Quota	Initial
Parks	49,875	3.5538	4
King	166,158	11.8395	12
Baker	82,888	5.9061	6
Bates	626,667	44.6528	45
Wilkins	126,979	9.0478	9
Total	1,052,567		76

This is one too many, so we need to increase the divisor. Let's try 14,100:

	÷14,100		
County	Population	Quota	Initial
Parks	49,875	3.5372	4
King	166,158	11.7843	12
Baker	82,888	5.8786	6
Bates	626,667	44.4445	44
Wilkins	126,979	9.0056	9
Total	1,052,567		75

This gives us exactly 75, so we're done. This is how it would look all in one table:

	÷14,034.22667			*÷14,100*	
County	Population	Quota	Initial	2nd Quota	2nd Try
Parks	49,875	3.5538	4	3.5371	4
King	166,158	11.8395	12	11.7843	12
Baker	82,888	5.9061	6	5.8786	6
Bates	626,667	44.6528	45	44.4445	44
Wilkins	126,979	9.0478	9	9.0056	9
Total	1,052,567		76		75

□

Like Jefferson's method, Webster's method carries a bias in favor of larger states but rounding the quotas to the nearest whole number greatly reduces this bias. Notice that Bates County, the largest, is the one that gets a representative trimmed because of the increased quota.

Also, like Jefferson's method, Webster's method does not always follow the quota rule, but it follows the quota rule much more often than Jefferson's method does. In fact, if Webster's method had been applied to every apportionment of Congress in all of U.S. history, it would have followed the quota rule every single time.

5.1.8 Land Rights, Citizenship and Voting Rights

In 1830 the ***Indian Removal Act*** was signed by President Jackson. The Act relocated Native Americans to land west of the Mississippi and the forced journey is known as the ***Trail of Tears***. To learn more, here is a link to an online museum exhibit with a timeline of Native American history and artifacts[1].

Toward the end of the Civil War, slaves were freed by Abraham Lincoln's ***Emancipation Proclamation*** on January 1st, 1863, but it was not until June 19th, 1865 that the last group of slaves in Texas learned they had been freed. This is now celebrated as Juneteenth[2]. You can read more about the timeline of slavery in this Jim Crow Museum of Racist Memorabilia[3].

Although they were technically free, Black people still faced unequal and unjust treatment. After Black men were granted the right to vote in 1869, ***voter suppression*** laws like poll taxes and literacy tests were implemented barriers to voting (American Civil Liberties Union, 2020).

In 1887, the ***Dawes Act*** was passed, which allotted Native American land to individuals and 60 million acres were taken by the government or non-Indian homesteaders (Indian Land Trust Tenure Foundation, 2020). This act also provided citizenship under certain restrictions which theoretically allowed native men to vote. All women were granted the right to vote in 1920, but the barriers enacted for men of color applied to women of color as well. You can read more about voting rights and suppression in this online timeline[4].

5.1.9 Huntington-Hill Method

In 1920, no new apportionment was done, because Congress couldn't agree on the method to be used. They appointed a committee of mathematicians to investigate, and they recommended the Huntington-Hill Method. They continued to use Webster's method in 1931, but after a second report recommending Huntington-Hill, it was adopted in 1941 and is the method of apportionment still used today.

The Huntington-Hill Method is similar to Webster's method, but attempts to minimize the difference in the percentage of how many people each representative will represent.

Huntington-Hill Method.

The first two steps are the same as the previous methods

1. Determine how many people each representative should represent. Do this by dividing the total population of all the states by the total number of representatives. This answer is called the ***divisor***.

2. Divide each state's population by the divisor to determine how many representatives it should have. Record this answer to several decimal places. This answer is called the ***quota***.

[1] http://recordsofrights.org/themes/4/rights-of-native-americans
[2] https://www.juneteenth.com/history.htm
[3] https://www.ferris.edu/htmls/news/jimcrow/timeline/slavery.htm
[4] https://www.carnegie.org/topics/topic-articles/voting-rights/voting-rights-timeline/

3. Cut off the decimal part of the quota to obtain the ***lower quota***, which we'll call n. Compute $\sqrt{n(n+1)}$, which is the geometric mean of the lower quota and one value higher.
4. Instead of using 0.5 to round like we are used to, we use the geometric mean. If the quota is as large or larger than the geometric mean, round up; if the quota is smaller than the geometric mean, round down. This is the ***initial*** allocation. Add up the resulting whole numbers.
5. If the total from Step 4 is less than the total number of representatives, reduce the divisor and recalculate the quota and allocation. If the total from step 4 is larger than the total number of representatives, increase the divisor and recalculate. Continue doing this until the total is equal to the exact number of representatives. The divisor we end up using is called the ***modified divisor***. This is a trial-and-error process.

5.1.10 What is the Geometric Mean?

Let's look at why Hill and Huntington decided to use the geometric mean for rounding instead of the arithmetic mean of 0.5, which is halfway between two numbers. By calculating some geometric means and looking at them in a table we can see a pattern:

	Geometric Mean
n	$\sqrt{n(n+1)}$
1	$\sqrt{(1(1+1)} = \sqrt{1(2)} \approx 1.414$
2	$\sqrt{(2(2+1)} = \sqrt{2(3)} \approx 2.449$
3	$\sqrt{(3(3+1)} = \sqrt{3(4)} \approx 3.464$
4	$\sqrt{(4(4+1)} = \sqrt{4(5)} \approx 4.472$
...	
10	$\sqrt{(10(10+1)} = \sqrt{10(11)} \approx 10.488$
...	
100	$\sqrt{(100(100+1)} = \sqrt{100(101)} \approx 100.499$

Notice that each geometric mean is between n and $n+1$. For smaller numbers, the decimal part is less than 0.5 and as n gets larger, the decimal part gets closer and closer to 0.5. This gives smaller states a chance to round up before larger states. That's how the Hill-Huntington method reduces the bias in favor of larger states.

Now let's see how to use the geometric mean in an example.

Example 5.1.7 Again, we will practice with Floyd, with an initial divisor of 21,900.82927 and 41 representatives:

County	Population	Quota	Lower Quota	Geometric Mean	Initial
King	162,310	7.4111	7	$\sqrt{7(8)} \approx 7.48$	7
Garner	538,479	24.5872	24	$\sqrt{24(25)} \approx 24.49$	25
Taylor	197,945	9.0017	9	$\sqrt{9(10)} \approx 9.49$	9
Total	897,934				41

Notice that for Garner, the quota of 24.5872 is above the geometric mean of 24.49, so we round up to 25. This gives the required total, so we're done. □

Example 5.1.8 Again, here is Lewis, with an initial divisor of 14,034.22667:

	÷14,034.22667				
County	Population	Quota	Lower Quota	Geometric Mean	Initial
Parks	49,875	3.5538	3	$\sqrt{3(4)} \approx 3.46$	4
King	166,158	11.8395	11	$\sqrt{11(12)} \approx 11.49$	12
Baker	82,888	5.9061	5	$\sqrt{5(6)} \approx 5.48$	5
Bates	626,667	44.6528	44	$\sqrt{44(45)} \approx 44.50$	45
Wilkins	126,979	9.0478	9	$\sqrt{9(10)} \approx 9.49$	9
Total	1,052,567				76

We end up with 76 which is too many, so we need to increase the divisor. Let's try 14,100:

	÷14,100				
County	Population	Quota	Lower Quota	Geometric Mean	Initial
Parks	49,875	3.5372	3	≈ 3.46	4
King	166,158	11.7843	11	≈ 11.49	12
Baker	82,888	5.8786	5	≈ 5.48	5
Bates	626,667	44.4445	44	≈ 44.50	45
Wilkins	126,979	9.0056	9	≈ 9.49	9
Total	1,052,567				75

This time Bates had a quota less than its geometric mean, so it did not get rounded up and we have exactly 75 representatives.

In both these cases, the apportionment produced by the Huntington-Hill method was the same as those from Webster's method, but that will not always be the case. □

In 1980, two mathematicians, Peyton Young and Mike Balinski, proved what we now call the Balinski-Young Impossibility Theorem.

Balinski-Young Impossibility Theorem.

The Balinski-Young Impossibility Theorem shows that any apportionment method which always follows the quota rule will be subject to the possibility of paradoxes like the Alabama, New States, or Population paradoxes. In other words, we can choose a method that avoids those paradoxes, but only if we are willing to give up the guarantee of following the quota rule.

Example 5.1.9 Consider a small country with 5 states, two of which are much larger than the others. We need to apportion 70 representatives. Apportion the representatives using both Webster's method and the Huntington-Hill method.

State	Population
A	300,500
B	200,000
C	50,000
D	38,000
E	21,500

Solution. Step 1: The total population is 610,000. Dividing this by the 70 representatives gives the divisor: 8714.286.

Step 2: Dividing each state's population by the divisor gives the quotas.

State	Population	Quota	Initial
A	300,500	34.48361	34
B	200,000	22.95082	6
C	50,000	5.737705	6
D	38,000	4.360656	4
E	21,500	2.467213	2

Webster's Method

Step 3: Using Webster's method, we round each quota to the nearest whole number using the rounding rule of 0.5 or higher to round up.

State	Population	Quota	Initial
A	300,500	34.48361	34
B	200,000	22.95082	23
C	50,000	5.737705	6
D	38,000	4.360656	4
E	21,500	2.467213	2

Step 4: Adding these up only gives us 69 representatives, so we adjust the divisor down. We try 8,700, which gives us 70 representatives. Notice that State A, the largest state, is the one that got rounded up the second time.

	÷8,700		
State	Population	Quota	Initial
A	300,500	34.54023	35
B	200,000	22.98851	23
C	50,000	5.747126	6
D	38,000	4.367816	4
E	21,500	2.471264	2

Huntington-Hill Method

Step 3: Using the Huntington-Hill method, we cut off the decimal to find the lower quota, then calculate the geometric mean based on each lower quota. If the quota is less than the geometric mean, we round down; if the quota is more than the geometric mean, we round up.

State	Population	Quota	Lower Quota	Geometric Mean	Initial
A	300,500	34.48361	34	34.49638	34
B	200,000	22.95082	23	22.49444	23
C	50,000	5.737705	6	5.477226	6
D	38,000	4.360656	4	4.472136	4
E	21,500	2.467213	2	2.44949	3

These allocations add up to 70. Notice that this allocation is different than that produced by Webster's method. In this case, state E, which is smaller, got one more seat and state A got one less. □

In this section we have learned four different methods of apportionment used in U.S. history, intertwined with the U.S. treatment of women, Native Americans and Black people. In the next section we will continue with our brief history of voting rights and look at different methods used for voting.

5.1.11 Exercises

In exercises 1-14, determine the apportionment using

a. Hamilton's Method

b. Jefferson's Method

c. Webster's Method

d. Huntington-Hill Method

1. A college offers tutoring in Math, English, Chemistry, and Biology. The number of students enrolled in each subject is listed below. If the college can only afford to hire 15 tutors, determine how many tutors should be assigned to each subject.

Math: 330	English: 265	Chemistry: 130	Biology: 70

2. Reapportion the previous problem if the college can hire 20 tutors.

3. The number of salespeople assigned to work during a shift is apportioned based on the average number of customers during that shift. Apportion 20 salespeople given the information below.

Shift	Morning	Midday	Afternoon	Evening
Avereage number of customers	95	305	435	515

4. Reapportion the previous problem if the store has 25 salespeople.

5. Three people invest in a treasure dive, each investing the amount listed below. The dive results in 36 gold coins. Apportion those coins to the investor.

Aisha: \$7,600	Basir: \$5,900	Carlos: \$1,400

6. Reapportion the previous problem if 37 gold coins are recovered.

7. A small country consists of four states, whose populations are listed below. If the legislature has 116 seats, apportion the seats.

A: 33,700	B: 559,500	C: 141,300	D: 89,100

8. Reapportion the previous problem with 124 seats.

9. A small country consists of five states, whose populations are listed below. If the legislature has 119 seats, apportion the seats.

A: 810,000	B: 473,000	C: 292,000	D: 594,000	E: 211,000

10. Reapportion the previous problem with 126 seats.

11. A small country consists of six states, whose populations are listed below. If the legislature has 200 seats, apportion the seats.

A: 3,411	B: 2,421	C: 11,586	D: 4,494	E: 3,126	F: 4,962

12. Reapportion the previous problem with 180 seats

13. A small country consists of six states, whose populations are listed below. If the legislature has 250 seats, apportion the seats.

A: 82,500	B: 84,600	C: 96,000	D: 98,000	E: 356,500	F: 382,500

14. Reapportion the previous problem with 240 seats.

15. A small country consists of three states, whose populations are listed below.

A: 6,000	B: 6,000	C: 2,000

a. If the legislature has 10 seats, use Hamilton's method to apportion the seats.

b. If the legislature grows to 11 seats, use Hamilton's method to apportion the seats.

c. Explain what happened in part b. What do you think would be a fair solution?

d. Try Jefferson's method for 11 seats. Does that solve the problem?

16. A small country consists of three states, whose populations are listed below.

A: 10,000	B: 10,000	C: 1,000

a. If the legislature has 10 seats, use Hamilton's method to apportion the seats. Which rule is not met in this case?

b. If the legislature grows to 11 seats, use Hamilton's method to apportion the seats.

c. If there could only be 10 seats, what do you think would be a fair solution?

17. A state with five counties has 50 seats in their legislature. Using Hamilton's method, apportion the seats based on the 2010 census, then again using the 2020 census. Explain what happened in 2020 apportionment. Do you think it is fair?

County	2000 Population	2010 Population
Douglass	60,000	60,000
Parks	31,200	31,200
King	69,200	72,400
Du Bois	81,600	81,600
Lewis	118,000	118,400

18. A state with five counties has 62 seats in their legislature. Using Hamilton's method, apportion the seats based on the 2010 census, then again using the 2010 census. Explain what happened in 2020 apportionment. Do you think it is fair?

County	2000 Population	2010 Population
Gray	75,000	83,200
Castile	89,000	89,000
Brown	32,500	32,500
Taylor	153,000	153,000
Floyd	109,000	112,000

19. A school district has two high schools: Clatsop, serving 1715 students, and Siletz, serving 7364. The district could only afford to hire 13 guidance counselors.

a. Determine how many counselors should be assigned to each school using Hamilton's method.

b. The following year, the district expands to include a third school, Cayuse, serving an additional 2989 students. Based on the divisor from above, how many additional counselors should be hired for Cayuse?

c. After hiring that many new counselors, the district recalculates the reapportion using Hamilton's method. Determine the outcome.

d. Explain what happened in the new apportionment. Do you think the outcome from part c is fair? Why or why not.

20. A school district has two middle schools: Tubman, serving 451 students, and Blackshear, serving 176. The district could afford to hire 8 art teachers.

a. Determine how many art teachers should be assigned to each school using Hamilton's method.

b. The following year, the district expands to include a third school, Banneker, serving 231 additional students. Based on the divisor from above, how many additional art teachers should be hired for the new school?

c. After hiring that many new art teachers, the district recalculates the reapportion using Hamilton's method. Determine the outcome.

d. Explain what happened in the new apportionment. Do you think the outcome from part c is fair? Why or why not.

5.2 Voting Methods

Objectives: Section 5.2

Students will be able to:

- Read a voter preference schedule for ranked choice voting
- Calculate the minimum number of votes to win a majority
- Find the winner of an election using the plurality method
- Find the winner of an election using the instant runoff method
- Find the winner of an election using the Borda count method
- Find the winner of an election using the pairwise (Condorcet) method

5.2.1 Voting Rights

Now that we have studied how to apportion the number of representatives for each state, along with a brief history of the U.S. government, we want to look at how representatives and other public officials are elected. There are also many other situations where a group consensus is needed, and some type of voting occurs. You may be most familiar with systems where you get one vote and the candidate with the most votes wins. However, there are many other ways of determining a winner. We will look at several voting methods in this section.

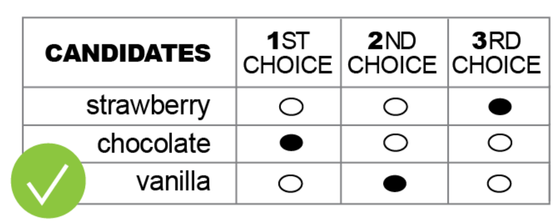
For an election using any of these methods to be fair, each person must have an equal right to vote and an equal ability to exercise that right. The ***Voting Rights Act of 1965*** prohibited any election practice that denies the right to vote based on race, but key portions of the law have been invalidated since then. ***Voter suppression*** continues today in different forms such as purging voter rolls, eliminating polling places and voter identification laws, which can also affect transgender people.

The John Lewis Voting Rights Advancement Act was passed in the U.S. House of Representatives in 2019 to restore elements of the Voting Rights Act and make it possible to enforce the law. It has not been taken up in the Senate, though. Follow this link if you would like to read a more detailed timeline of voting rights[1].

5.2.2 Ranked Choice Ballots

To use some of the methods we are going to study, we need to know more than just each person's first choice. We also need to know their 2nd and 3rd choices, and so on. A ballot where each person ranks all of the choices in order of their preference is called a ***ranked choice ballot***.

The image to the right shows an example of a ranked choice ballot[2]. Ranked choice ballots can be used in elections by mail or in person at polling places. Let's see how to tally the results from a ranked choice election.



[1] https://www.aclu.org/voting-rights-act-major-dates-history
[2] "Ranked Choice Ballot", Copyright 2020 Ranked Choice Voting Resource Center is licensed under CC BY-SA 3.0

Example 5.2.1 Students are voting for their class president and the candidates are Omar (O), Ana (A), and Helena (H). They have ranked the candidates according to their preference.

	Vien	Ana	Marv	Tasha	Eve	Omar	Lupe	Dave	Helena	Jimmy
1st choice	A	A	O	H	A	O	H	O	H	A
2nd choice	O	H	H	A	H	H	A	H	A	H
3rd choice	H	O	A	O	O	A	O	A	O	O

□

5.2.3 Preference Schedule

To simplify this data, we create a ***preference schedule*** which combines all the people who voted with the exact same ranking. For example, Tasha, Lupe, and Helena, all voted in the same way, HAO. So, we will list that combination once, with the number 3 at the top of that column. Three students also voted for AHO, and OHA. One student voted for the ranking AOH. Below is the preference schedule.

	1	3	3	3
1st choice	A	A	O	H
2nd choice	O	H	H	A
3rd choice	H	O	A	O

Notice that by totaling the vote counts across the top of the preference schedule we can see the total number of votes cast: $1+3+3+3=10$ total votes.

5.2.4 Plurality Method

The voting method you may be most familiar with in the United States is the ***plurality method***. In the plurality method, the candidate with the most first-choice votes is declared the winner.

There is a difference between a ***majority*** and a plurality. For a majority, a candidate must have over 50% of the votes. There are some states that require a majority in order to win an election. In more cases, though, only a plurality is required, which means having more votes than any other candidate. If a tie occurs without using ranked choice voting, then a new run-off election would be required.

Example 5.2.2 In our election from above, we had this preference table:

	1	3	3	3
1st choice	A	A	O	H
2nd choice	O	H	H	A
3rd choice	H	O	A	O

For the plurality method, we only look at the 1st choice row. Totaling them up:

- Ana: $1+3=4$ votes
- Omar: 3 votes
- Helena: 3 votes

Ana is the winner using the plurality voting method.

Notice that Ana won with 4 out of 10 votes, or 40% of the votes, which is a plurality of the votes, but not a majority. □

5.2.5 How Many Votes are Needed to Win?

If we know how many voters there are, we can calculate the minimum number of votes needed for a majority win.

Continuing Example 5.2.2 from above, there are 10 voters, so we divide that by 2 or multiply by 0.5 to find 50%. Then we round up to the next whole number to get just over 50%.

$10 \div 2 = 5$. Then we round up to 6 votes.

Exactly half of the votes would be a tie, so we round up to the next whole number. A candidate would need 6 out of 10 votes to win a majority. If there was an odd number of voters we would divide by 2 and get a decimal. Let's say there were 13 voters:

$13 \div 2 = 6.5$. Then we round up to 7 votes.

Then we round up to the next whole number, which is 7 votes. A candidate can win a plurality with fewer votes than the majority, as long as they have more than any other candidate.

5.2.6 Insincere Voting

In a system with only two major political parties like the United States, it can be nearly impossible for third-party candidates to break through. You may feel like you would be wasting your vote if you voted for a third-party candidate. Consider this example.

Example 5.2.3 Here is a two-party election with preferences shown below. It is a close race, but candidate B would win.

Number of voters	96	100
1st choice	A	B
2nd choice	B	A

Suppose a third candidate, C, entered the race, and a segment of voters sincerely want to vote for that third candidate. Here is the new preference schedule.

Number of voters	96	90	10
1st choice	A	B	C
2nd choice	B	A	B
3rd choice	C	C	A

The ten people who prefer candidate C are in a bind because if they vote for C, then A will win, but they prefer B over A. It is likely that they will vote for B, their second choice, to keep A from winning. ☐

Situations like this can lead to insincere voting. ***Insincere voting*** is when a person doesn't vote according to their actual preference for strategic purposes. In the case above, a person who wants to vote for a third-party candidate realizes it would take a vote away from their major party candidate. Not wanting to see their party lose the election, they vote for the major party candidate.

5.2.7 Instant Runoff Method

The ***Instant Runoff Method*** is a modification of the plurality method that attempts to address the issue of insincere voting.

The voting is done with ranked choice ballots and a preference schedule is generated. Then we tally all the first-choice votes. The candidate with the *least* first-place votes is eliminated, but any votes for that candidate are transferred to the voters' next choice. The process continues until one candidate has a majority. We will show you how it works with an example.

Example 5.2.4 Consider the preference schedule below, in which a committee is choosing a chairperson. The candidates are labeled A, B, C, D, and E for simplicity. Here is the preference schedule:

	3	5	1	5	7	2	1
1st choice	B	C	C	B	D	A	E
2nd choice	C	A	B	D	C	E	D
3rd choice	A	E	B	C	A	B	B
4th choice	D	D	E	A	E	C	A
5th choice	E	B	A	E	B	D	C

Let's start by tallying the first-choice

$$\text{A: } 2 \quad \text{B: } 3+5=8 \quad \text{C: } 5+1=6 \quad \text{D: } 7 \quad \text{E: } 1$$

If this was a plurality election, B would be the winner with 8 first-choice votes.

Round 1: We make our first elimination. Choice E has the fewest first-place votes, so we eliminate E from the election. To transfer the 1 vote, we look at the person who voted for E in the last column. Their second choice was D, so we transfer that one vote to D, who will now have 8 votes:

$$\begin{array}{ccccc} \text{A: } 2 & \text{B: } 8 & \text{C: } 6 & \text{D: } 7 & \cancel{\text{E: } 1} \\ & & & \underline{+1} & \\ & & & 8 & \end{array}$$

Round 2: There are 24 votes in total, so a majority would be 13 votes. Since no candidate has the majority, we eliminate the next candidate with the fewest votes, which is A. We look at the column for the two people who chose A. Their second choice was E, who is already eliminated. Their third choice is B, so their 2 votes are transferred to B, for a total of 10:

$$\begin{array}{ccccc} \cancel{\text{A: } 2} & \text{B: } 8 & \text{C: } 6 & \text{D: } 7 & \cancel{\text{E: } 1} \\ & \underline{+2} & & \underline{+1} & \\ & 10 & & 8 & \end{array}$$

Round 3: Now C has the fewest votes, and the there are two columns for C in the preference schedule. Five people who voted for C chose A second, then E (who have already been eliminated), then D. So, we transfer their 5 votes to D. One person who voted for C selected B as their second choice, so their one vote goes to B. Then we tally the votes for each remaining candidate.

$$\begin{array}{ccccc} \cancel{\text{A: } 2} & \text{B: } 8 & \cancel{\text{C: } 6} & \text{D: } 7 & \cancel{\text{E: } 1} \\ & \underline{+2} & & \underline{+1} & \\ & 10 & & 8 & \\ & \underline{+1} & & \underline{+5} & \\ & 11 & & 13 & \end{array}$$

Now we can see that D is the winner of the instant runoff method with 13 votes. We can stop this method earlier if a candidate reaches a majority. □

The Instant Runoff Method is similar to the idea of holding runoff elections, but since every voter's order of preference is recorded on the ballot, the runoff can be computed quickly without requiring a second costly election. It also makes it safe to vote for a third-party candidate knowing that your vote will go to your second choice if your first choice can't win.

This voting method is used in several places around the world, including the election of members of the Australian House of Representatives, statewide elections in Maine, and local elections in Benton County, Oregon (Ranked Choice Voting Resource Center, 2020).

5.2.8 Borda Count (Point System)

Borda Count is another voting method, named for Jean-Charles de Borda, who developed the system in 1770. In this method, points are assigned to each candidate based on their ranking; 1 point for last choice, 2 points for second-to-last choice, and so on. The point values are totaled, and the candidate with the largest point total is the winner. We'll show you how to do it in the next example.

Example 5.2.5 A group of PCC students are getting together for a student leadership conference and they need to decide where to meet. Their members are from four campuses: Sylvania, Rock Creek, Cascade and Southeast. The votes for where to hold the conference were:

	31	25	10	14
1st choice	Sylvania	Rock Creek	Southeast	Cascade
2nd choice	Southeast	Cascade	Cascade	Rock Creek
3rd choice	Rock Creek	Sylvania	Sylvania	Southeast
4th choice	Cascade	Southeast	Rock Creek	Sylvania

Now we will add a column on the left side of the table for the points. Then we multiply the points by the number of voters who put each candidate in that rank,

Points		31	25	10	14
4	1st choice	Sylvania	Rock Creek	Southeast	Cascade
3	2nd choice	Southeast	Cascade	Cascade	Rock Creek
2	3rd choice	Rock Creek	Sylvania	Sylvania	Southeast
1	4th choice	Cascade	Southeast	Rock Creek	Sylvania

We will start with Sylvania in the bottom row and move up. They have one point times 14 voters, 2 points times 25 + 10 voters, 3 points times zero, and 4 points times 31 voters. Multiplying and then adding gives us a total of 208 points for Sylvania. You can see the calculation below and we will do the same thing for each campus:

- Sylvania: $1 \cdot 14 + 2 \cdot 35 + 3 \cdot 0 + 4 \cdot 31 = 208$ points
- Rock Creek: $1 \cdot 10 + 2 \cdot 31 + 3 \cdot 14 + 4 \cdot 25 = 214$ points
- Cascade: $1 \cdot 31 + 2 \cdot 0 + 3 \cdot 35 + 4 \cdot 14 = 192$ points
- Southwest: $1 \cdot 25 + 2 \cdot 14 + 3 \cdot 31 + 4 \cdot 10 = 186$ points

Using the Borda count method, Rock Creek is the winning location. □

Note that Sylvania would have won with the plurality method. Borda count is sometimes described as a consensus-based voting system, since it can be used to choose a more broadly acceptable option over the one with the most first-choice support. This is a different approach than plurality and instant runoff voting that focus on first-choice votes; Borda Count considers every voter's entire ranking to determine the outcome.

Because of this consensus behavior, Borda Count, or some variation of it, is commonly used in awarding sports awards. Variations are used to determine the Most Valuable Player in baseball, to rank teams in NCAA sports, and to award the Heisman trophy.

5.2.9 Pairwise Comparison: Copeland's Method

Another method, called ***pairwise comparison***, the ***Condorcet method*** or ***Copeland's method***, attempts to be fair by looking at each pair of candidates separately.

In this method, we look at each pair as if they were the only two candidates running and determine which of the two is more preferred. In Copeland's method, the more preferred candidate is awarded 1 point. If there is a tie, each candidate is awarded $\frac{1}{2}$ point. After all the pairwise comparisons are made, the candidate with the most points, and hence the most pairwise wins, is declared the winner.

Variations of Copeland's Method are used in many professional organizations, including election of the Board of Trustees for the Wikimedia Foundation that runs Wikipedia.

Example 5.2.6 Consider our class president example from the beginning of the chapter. Determine the winner using Copeland's Method.

	1	3	3	3
1st choice	A	A	O	H
2nd choice	O	H	H	A
3rd choice	H	O	A	O

We need to look at each pair of candidates to see which would win in a one-to-one comparison. First, we will systematically list all of the pairs of candidates:

- Helena vs. Omar
- Helena vs. Ana
- Ana vs. Omar

Next, comparing Helena with Omar, we see that 6 voters, those marked in bold in the table below, would prefer Helena over Omar. Note that Helena doesn't have to be the voter's first choice – we're imagining that Ana wasn't in the race. The other 4 people chose Omar over Helena.

	1	***3***	3	***3***
1st choice	A	A	O	***H***
2nd choice	O	***H***	H	A
3rd choice	H	***O***	A	***O***

Based on this comparison of Helena vs. Omar, Helena wins 1 point.

Next, comparing Helena with Ana, the 6 voters in the last two columns prefer Helena to Ana, so Helena gets one point. Lastly, we compare Ana with Omar. The 1 voter in the first column prefers Ana, as do the 3 voters in the second column. The 3 voters in the third column prefer Omar, but the 3 voters in the last column would choose Ana. So, altogether $1 + 3 + 3 = 7$ voters prefer Ana over Omar. Ana gets 1 point.

To summarize, we write how many voters prefer each candidate ignoring all other candidates:

- Helena-6 vs Omar-4 (Helena gets 1 point)
- Helena-6 vs Ana-4 (Helena gets 1 point)
- Ana-7 vs Omar-3 (Ana gets 1 point)

Ana has a total of 1 point and Helena has 2 points, so Helena is the winner under Copeland's Method. This method gave us a different winner than the plurality method where Ana won, because more people preferred Helena to Ana, even though they didn't choose Helena for their first choice. □

Now that we've seen an example of each voting method, let's run through each method for the same scenario.

Example 5.2.7 Consider an advertising team voting to choose one of four different slogans, labeled A, B, C and D. Determine the winner using each method we have learned.

	5	3	6	4	2
1st choice	D	A	C	B	A
2nd choice	A	C	B	D	D
3rd choice	C	B	A	A	C
4th choice	B	D	D	C	B

Solution. ***Plurality Method:*** We tally the first-place votes:

$$\text{A: } 3+2=5 \quad \text{B: } 4 \quad \text{C: } 6 \quad \text{D: } 5$$

C wins the plurality method.

Instant Runoff Method: We start with the plurality tallies. Then we look for the slogan with the least first-choice votes, and that is B. Looking in the preference schedule at the 4 people who voted for B, we see their second choice is D. So we transfer those 4 votes to D, which now has 9.

$$\begin{array}{cccc} \text{A: } 5 & \cancel{\text{B: } 4} & \text{C: } 6 & \text{D: } 5 \\ & & & \underline{+4} \\ & & & 9 \end{array}$$

Now A has the least number of votes with 5, so we look at the voters who chose A for their first choice. There are two columns, so three of those votes will be transferred to C and 2 will be transferred to D.

$$\begin{array}{cccc} \cancel{\text{A: } 5} & \cancel{\text{B: } 4} & \text{C: } 6 & \text{D: } 5 \\ & & & \underline{+4} \\ & & & 9 \\ & & \underline{+3} & \underline{+2} \\ & & 9 & 11 \end{array}$$

D is the winner of the instant runoff method.

Borda Count: We add a column on the left for the points, starting with 1 point for last place and counting up. Then we multiply the points by the number of votes and add them all up.

Points		5	3	6	4	2
4	1st choice	D	A	C	B	A
3	2nd choice	A	C	B	D	D
2	3rd choice	C	B	A	A	C
1	4th choice	B	D	D	C	B

- A: $1 \cdot 0 + 2 \cdot 10 + 3 \cdot 5 + 4 \cdot 5 = 55$ points
- B: $1 \cdot 7 + 2 \cdot 3 + 3 \cdot 6 + 4 \cdot 4 = 47$ points
- C: $1 \cdot 4 + 2 \cdot 7 + 3 \cdot 3 + 4 \cdot 6 = 51$ points
- D: $1 \cdot 9 + 2 \cdot 0 + 3 \cdot 6 + 4 \cdot 5 = 47$ points

A is the winner of the Borda count method.

Pairwise Comparison:

First we list all the possible pairs:

- A vs B
- A vs C
- A vs D
- B vs C
- B vs D
- C vs D

Then we count the number of voters who would choose each candidate if they were the only two in the race.

- A-10 vs B-10
- A-14 vs C-6
- A-11 vs D-9
- B–4 vs C-16
- B-13 vs D-7
- C-9 vs D-11

Totaling one point for each win and half a point for each tie gives us:

- A has 2.5 points
- B has 1.5 points
- C has 1 point
- D has 1 point

A is the winner of the pairwise comparison method. □

5.2.10 Which Method is the Most Fair?

To see a very simple example of how difficult voting can be, consider the election below:

	5	5	5
1st choice	A	C	B
2nd choice	B	A	C
3rd choice	C	B	A

Notice that in this election: 10 people prefer A to B, 10 people prefer B to C, and 10 people prefer C to A.

No matter whom we choose as the winner, $\frac{2}{3}$ of voters would prefer someone else! This scenario is called ***Condorcet's Voting Paradox***. In this situation, there is no fair resolution. There are many additional paradoxes and fairness criteria that are explained in David Lippman's original version of *Math in Society*[3].

Since it is not possible to have one method that is fair in every situation, it is important to decide which method is the most fair for any given situation.

5.2.11 Primaries and Sequential Voting

For many public offices in the U.S., a sequence of two public votes is held: a primary election and the general election. For non-partisan offices like sheriff and judge, in which political party affiliation is not declared, the primary election is usually used to narrow the field of candidates.

For partisan positions, typically, either the top two candidates receiving the most votes move forward, or the top candidate from each party will move forward. This is called ***sequential voting*** - a process in which voters cast totally new ballots after each round of eliminations. Sequential voting has become quite common in television, where it is used in reality shows like *The Voice*.

In some states a ***closed primary*** is used, in which only voters of each party can vote for that party's

[3]Math in Society: Voting Theory, by David Lippman

candidates. In other states, an ***open primary*** is used, where voters can pick the party whose primary they want to vote in. In other states, ***caucuses*** are used, which are meetings of the political parties, only open to party members. Closed primaries are often disliked by independent voters, who like the flexibility to change which party they are voting in. Open primaries have the disadvantage of raiding, though, where a voter could vote in their non-preferred party's primary with the intent of selecting a weaker opponent for their preferred party's candidate.

Regardless of the primary type, the general election is the main election. While rules vary state-to-state, for an independent or minor party candidate to get listed on the ballot, they typically have to gather a certain number of signatures to petition for inclusion.

In the next section we will look at a different type of election for the President of the United States.

5.2.12 Exercises

1. To decide on a new website design, the designer asks people to rank three designs that have been created (labeled A, B, and C). The individual ballots are shown below. Create a preference table.

ABC, ABC, ACB, BAC, BCA, BCA, ACB, CAB, CAB, BCA, ACB, ABC

2. To decide on a movie to watch, a group of friends all vote for one of the choices (labeled A, B, and C). The individual ballots are shown below. Create a preference table.

CAB, CBA, BAC, BCA, CBA, ABC, ABC, CBA, BCA, CAB, CAB, BAC

In exercises 3-18, complete the following:

a. How many voters voted in this election?

b. How many votes are needed for a majority?

c. Find the winner under the plurality method.

d. Find the winner under the Instant Runoff Voting method.

e. Find the winner under the Borda Count Method.

f. Find the winner under Copeland's method.

3. The planning committee for a renewable energy trade show is trying to decide where to hold their next trade show in. The votes are shown below.

Number of voters	9	19	11	8
1st choice	Buffalo	Atlanta	Chicago	Buffalo
2nd choice	Atlanta	Buffalo	Buffalo	Chicago
3rd choice	Chigaco	Chicago	Atlanta	Atlanta

4. A non-profit agency is electing a new chair of the board. The votes are shown below.

Number of voters	11	5	10	4	3
1st choice	Abdulla	Cantos	Beck	Cantos	Abdulla
2nd choice	Cantos	Beck	Cantos	Abdulla	Beck
3rd choice	Beck	Abdulla	Abdulla	Beck	Cantos

5. A group of students is deciding which class to take together. The votes are shown below.

Number of voters	2	1	2	3	2	2
1st choice	Art	Art	Biology	Biology	Calculus	Calculus
2nd choice	Biology	Calculus	Art	Calculus	Art	Biology
3rd choice	Calculus	Biology	Calculus	Art	Biology	Art

6. A professional psychology association is deciding where to hold their next conference. The votes are shown below.

Number of voters	3	2	7	3	6	5
1st choice	Alaska	Alaska	Barbados	Barbados	California	California
2nd choice	Barbados	California	Alaska	California	Alaska	Barbados
3rd choice	California	Barbados	California	Alaska	Barbados	Alaska

7. The Hillsboro City Council has an election with three candidates, D, E, F. The votes are:

Number of voters	5	4	3	7	2	10
1st choice	D	D	E	E	F	F
2nd choice	E	F	F	D	E	D
3rd choice	F	E	D	F	D	E

8. The Vancouver mayor's race has three candidates, G, H, I. The votes are:

Number of voters	12	6	14	3	15	1
ast choice	G	G	H	H	I	I
2nd choice	H	I	I	G	G	H
3rd choice	I	H	G	I	H	G

9. The Beaverton City Council has an election for one open position. There are four candidates (labeled A, B, C, and D for convenience). The preference schedule for the election is:

Number of voters	120	50	40	90	60	100
1st choice	C	B	D	A	A	D
2nd choice	D	C	A	C	D	B
3rd choice	B	A	B	B	C	A
4th choice	A	D	C	D	B	C

10. Portland City Council has an election for one open position. There are four candidates (labeled E, F, G, H for convenience). The preference schedule for the election is:

Number of voters	11	8	3	9	16	14	9
1st choice	E	H	H	G	E	F	F
2nd choice	F	G	E	F	G	H	G
3rd choice	H	F	G	H	H	E	E
4th choice	G	E	F	E	F	G	H

11. The Portland mayor's race has four candidates, I, J, K, L. The votes are:

Number of voters	21	18	9	14	5	12	11	2
1st choice	I	K	K	L	J	I	K	J
2nd choice	L	I	J	K	I	K	L	I
3rd choice	J	L	I	I	K	L	I	L
4th choice	K	J	L	J	L	J	J	K

12. A Reynolds School Board position has four candidates, M, N, O, P. The votes are:

Number of voters	16	2	12	15	34	22	32	13	11
1st choice	M	N	N	P	O	M	M	O	N
2nd choice	O	O	P	O	M	P	N	M	O
3rd choice	P	M	O	N	N	O	P	P	M
4th choice	N	P	M	M	P	N	O	N	P

13. The Gresham mayor's race has four candidates, Q, R, S, T. The votes are:

Number of voters	11	4	15	18	14	5	13	10
1st choice	Q	S	Q	T	R	S	S	R
2nd choice	R	T	S	S	Q	Q	Q	T
3rd choice	S	R	T	Q	T	T	R	S
4th choice	T	Q	R	R	S	R	T	Q

14. The Parkrose School Board has an election for an open position with four candidates, U, V, W, X. The votes are:

Number of voters	16	21	15	5	3	11	9	8	22
1st choice	U	W	X	V	X	X	W	U	V
2nd choice	X	X	V	X	W	U	V	V	W
3rd choice	V	U	U	W	V	V	X	X	U
4th choice	W	V	W	U	U	W	U	W	X

15. A Multnomah county commissioner's race has five candidates, A, B, C, D, E. The votes are:

Number of voters	12	14	21	18	19	23
1st choice	B	C	E	E	A	B
2nd choice	A	A	B	A	C	D
3rd choice	D	E	A	B	B	C
4th choice	E	B	C	C	D	E
5th choice	C	D	D	D	E	A

16. The Oregon State governor's race has five candidates, F, G, H, I, J. The votes are:

Number of voters	16	12	9	21	17	11	6
1st choice	F	F	H	I	G	G	J
2nd choice	I	G	I	F	H	J	F
3rd choice	G	H	G	J	I	F	H
4th choice	J	J	J	H	F	I	G
5th choice	H	I	F	G	J	H	I

17. A Washington county commissioner's race has five candidates, K, L, M, N, O. The votes are:

Number of voters	22	19	13	26	14	5	17	11
1st choice	K	M	K	O	L	L	N	N
2nd choice	M	L	L	K	N	K	O	L
3rd choice	L	N	N	N	M	O	K	K
4th choice	N	O	M	L	K	M	M	O
5th choice	O	K	O	M	O	N	L	N

18. A Clackamas county commissioner's race has five candidates, P, Q, R, S, T. The votes are:

Number of voters	31	16	29	52	16	18	11	13	12
1st choice	R	R	P	Q	T	T	S	S	S
2nd choice	S	Q	R	S	Q	S	T	Q	P
3rd choice	T	T	T	R	S	Q	R	T	R
4th choice	P	S	Q	T	P	R	P	R	T
5th choice	Q	P	S	P	R	P	Q	P	Q

5.3 The Popular Vote, Electoral College and Electoral Power

Objectives: Section 5.3

Students will be able to:

- Calculate the number of electors per state
- Determine the winner of the popular vote
- Determine the winner of the electoral college
- Calculate the electoral power of each state

5.3.1 Choosing the U.S. President

In the last section we talked about equal voting rights and studied several voting methods. To choose the President of the United States, we use another method, called the ***Electoral College***. The Founders, with their newfound freedom from Britain, wanted to create a democracy where citizens could participate in choosing their president. Prior to the Electoral College, there were three options proposed (Best, 2004; Clayton, 2015). They were

1. Congress would elect the president
2. State Legislatures would vote and elect the president
3. The people would elect a president by popular vote

The first option of having Congress do this work was declined as this would alter the balance of power among the three separate branches of government. The idea to have State Legislatures also failed as there wouldn't be the separation between having an independent Federal Government and State Governments.

The ***popular vote*** means a candidate must win a plurality of all the votes cast for president, regardless of which state the voters live in. The popular vote was decided against as the Founders didn't think all citizens would be informed enough about the candidates to make an educated decision (Clayton, 2015).

Instead, the ***Electoral College*** was decided upon and, in this system, each state was given a certain number of electoral votes. Each state got an equal two votes for the two senate seats. This gave smaller states a leg up as they would have at least these votes. In addition, each state would be given more votes based on their population, equal to their representation in the House (Clayton, 2015). The District of Columbia, which is not a state, also gets 3 electoral votes. A candidate must win a majority of the electoral votes to win the U.S. presidency. If no candidate has a majority of electoral votes then a contingent election would be held by the U.S. House of Representatives between the three candidates with the most electoral votes.

The 4 million U.S. Citizens who live in the ***U.S. territories*** of Guam, the Virgin Islands, the Northern Mariana Islands and Puerto Rico cannot vote in the presidential elections and do not have any electors for the U.S. President. Additionally, there are U.S. Nationals in American Samoa and other territories that cannot vote. You can learn more about residents of these territories in this Census Bureau Report[1]. We will also reiterate that ***voter suppression*** still exists and elections will not be truly fair until all of its forms are eliminated.

As stated in the Constitution, the U.S. currently has 435 representatives in the House. They each serve 2-year terms, while the two senators from each state serve 6-year terms.

Example 5.3.1 How many electoral votes are there in the U.S. Electoral College?

[1] https://www.census.gov/prod/2011pubs/12statab/outlying.pdf

Solution. $100 + 435 + 3 = 538$

By adding all the U.S. Senators, Representatives and 3 for Washington, D.C., we have a total of 538 electoral votes. □

Example 5.3.2 How many electoral votes does a candidate need to win to become the President of the United States?

Solution. $538 \div 2 = 269$ then round up to 270.

A candidate needs 270 electoral votes to win the U.S. Presidency. □

Example 5.3.3 The number of people who voted in the presidential election in 2016 was 136,452,150 (Ballotpedia, n.d.). How many votes would be needed to win a majority of the popular vote?

Solution. $136,452,150 \div 2 = 68,226,075$ then round up to 68,226,076.

A candidate would need 68,226,076 votes to win the popular vote. □

Example 5.3.4 There were approximately 230,931,921 eligible voters at the time of the 2016 election (McDonald, 2016). What percentage of eligible voters voted?

Solution. $136,452,150 \div 230,931,921 = 0.591$

About 59% of all the eligible voters voted in the 2016 election. □

If you are interested in the 2016 election voting rates by race, Hispanic origin and age, you can find them in this Census Bureau Report[2].

5.3.2 How Electoral Votes are Determined

While each state has a set number of electoral votes based on the formula above, the federal government gave states the control of how they distribute and cast their electoral votes (Paiva, 2011). In 48 states, the candidate with the most votes wins all of that state's electoral votes, often called ***winner-takes-all***. In essence, a candidate could win with barely a majority of the popular vote in that state and the entire state's votes goes to that individual (Clayton, 2015). In 2 states, Maine and Nebraska, however, they use a district plan where instead of winner-takes-all, they use ***proportionality***. This means that they award electoral votes by how individual districts vote. Each district gives their electoral vote to the winner of their district and the remaining two votes go to the candidate that won statewide.

There is a movement to change from the Electoral College to the popular vote. One way this is happening is through the ***National Popular Vote Interstate Compact***[3]. States who adopt this compact agree to award their electoral votes to the winner of the national popular vote. If states and/or D.C. with at least 270 electoral votes adopt this compact, then the winner of the popular vote would win the electoral college and the presidency. At this time there are 196 electoral votes in states and D.C. who have enacted the bill and the bill has passed one or two legislative houses in additional states with 88 electoral votes.

After citizens cast their votes by mail or on election day on the first Tuesday in November, electors gather in December and are responsible for casting the votes for their state. Electors cannot be elected officials and are chosen by their parties. Electors are expected to vote in alignment with the state they represent, but there have been faithless electors in the past. The U.S. Supreme Court ruled in July of 2020 that states can penalize electors who do not vote according to the law of the state.

To see how the electoral college and popular votes work, we will look at a made-up country with a smaller number of states.

Example 5.3.5 Consider this country with 4 states. The rules for the number of senators and electors are the same as the U.S. government. Each state will get 1 representative for every 40,000 residents. For

[2]https://www.census.gov/newsroom/blogs/random-samplings/2017/05/voting_in_america.html
[3]https://www.nationalpopularvote.com/

simplicity we will ignore any remainders.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Small	154,600			
Medium	262,340			
Large	581,135			
Huge	1,362,070			

a. Determine the number of electors for each state and the total for the country.

b. How many electoral votes are needed to win the presidential election?

c. Determine which candidate wins the popular vote and the electoral college vote.

d. Write down all the possible combinations of states that would win the election based on the electoral college.

e. What is the fewest number of individual votes needed to win the Electoral College?

f. What is the smallest percentage of the population that you can win the Electoral College with?

Solution.

a. To find the number of electors for each state, we first divide each state population by 40,000 to determine the number of representatives they will have. We will drop the decimal remainders, similar to Hamilton's method.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Small	154,600	$154,600 \div 40,000 = 3.865 \to 3$		
Medium	262,340	$262,340 \div 40,000 = 6.559 \to 6$		
Large	581,135	$581,135 \div 40,000 = 14.528 \to 14$		
Huge	1,362,070	$862,070 \div 40,000 = 21.552 \to 21$		

State	Population	Number of Representatives	Number of Senators	Number of Electors
Small	154,600	3	2	
Medium	262,340	6	2	
Large	581,135	14	2	
Huge	1,362,070	21	2	

Then we will add the number of representatives and the senators for each state.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Small	154,600	3	2	5
Medium	262,340	6	2	8
Large	581,135	14	2	16
Huge	1,362,070	21	2	23
Totals	1,860,145	44	8	52

Now we can add up the total of all the electors from each state and we have:

$5 + 8 + 16 + 23 = 52$ electors.

b. To find out how many electoral votes are needed to win the presidential election, we calculate a majority of 52 electors.

$52 \div 2 = 26$, round up to 27

27 electoral votes are needed to win the presidential election in this state

c. To find the winner for each method, we need the results of an election. In our example, there are 2 candidates for the president, Candidate A and Candidate B. When a candidate wins in a state, they get all the electoral votes for that state. Given the votes for each candidate below, determine who wins the popular vote and who wins the electoral college vote.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Small	56,259	98,341		
Medium	130,082	132,258		
Large	278,177	302,958		
Huge	546,555	315,515		
Totals				

First, we will add up all the votes for each candidate in their columns. Next, we will look at each state individually to determine which candidate wins each state. Candidate B has more votes in Small, Medium and Large, so Candidate B gets all of the electoral votes in those states. Candidate A won Huge, so they get all of those electoral votes. We enter zero for the candidate who does not win in each state. Then we add up the electoral vote columns. Here is the completed table.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Small	56,259	98,341	0	5
Medium	130,082	132,258	0	8
Large	278,177	302,958	0	16
Huge	546,555	315,515	23	0
Totals	1,011,073	849,072	23	29

Now we can see that Candidate A wins the popular vote with 1,011,073 votes compared to 849,072 for Candidate B. But Candidate B wins the Electoral College with 29 electoral votes. So, Candidate B becomes the President of the United States.

There have been 5 U.S. presidential elections where the winner of the popular vote was not the winner of the electoral college. In all the other elections, the two methods agreed and would have elected the same candidate.

Try calculating the results if these were the votes instead:

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Small	56,259	98,341		
Medium	131,200	131,140		
Large	278,177	302,958		
Huge	546,555	315,515		
Totals				

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Small	56,259	98,341	0	5
Medium	131,200	131,140	8	0
Large	278,177	302,958	0	16
Huge	546,555	315,515	23	0
Totals	1,012,191	847,954	31	21

This time Candidate A won Medium and Huge. That gave them the electoral vote victory. Note that the race in Medium was very close, and we only changed 1,118 votes or 0.06% of the votes to Candidate A which gave them the Electoral College.

Opponents of the Electoral College system say that it causes candidates to focus on a few swing states, rather than campaigning equally in all states.

d. Let's look at which combinations of states the candidates could seek to win with.

Since 27 electoral votes are needed, we will list all the winning combinations starting with the smallest number of states to win. We see that a candidate must win either Huge or Large, because Small and Medium together are not enough votes to win

Combination	Small, 5	Medium, 8	Large, 16	Huge, 23	Total
Huge + Small	5			23	28
Huge + Medium		8		23	31
Huge + Large			16	23	39
Huge + Small + Medium	5	8		23	36
Huge + Medium + Large		8	16	23	47
Huge + Small + Large	5		16	23	44
Large + Small + Medium	5	8	16		29
All States	5	8	16	23	52

From the table we find there are 8 different ways to win the electoral college and candidates should definitely campaign in Large and Huge.

e. To find the fewest number of individual votes a candidate could win with; let's look at the table above. The combination of states with the fewest electoral votes is Huge and Small. Let's see how many votes would be needed to win those two states.

For Huge:

$862,070 \div 2 = 431,035$, then round up to 431,036

For Small:

$154,600 \div 2 = 77,300$, then round up to 77,301

$431,036 + 77,301$ votes $= 508,337$ votes

f. The smallest number of votes to win the electoral college is 508,337. What percentage of the population is that? We will divide by the total population to get

$508,337 \div 1,860,145 = 0.2733$

If a candidate had the right combination of states, they could win the Electoral College with only 27.3% of the vote.

People who support the Electoral College often say it protects small states. People who oppose the Electoral College often say that each person's vote should count equally as in the popular vote. Next we will look at electoral power to understand the effect on small states.

□

People who support the Electoral College often say it protects small states. People who oppose the Electoral College often say that each person's vote should count equally as in the popular vote. Next we will look at electoral power to understand the effect on small states.

5.3.3 Electoral Power

Electoral power is the value of a person's vote in one state compared to another state. A related topic is voting power, which is the probability of any one person influencing an election. If you are interested in voting power, you can look up the Banzhaf Power Index. To calculate the electoral power, we will calculate a ratio of electoral votes for a given number of people. We could use any number, so we will choose the number of people that each representative represents, which is our divisor of 40,000 people.

Example 5.3.6 Using the same fictional country, use the table below to calculate the electoral power of each state.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes Per 40,000 people
Small	154,600				
Medium	262,340				
Large	581,135				
Huge	1,362,070				

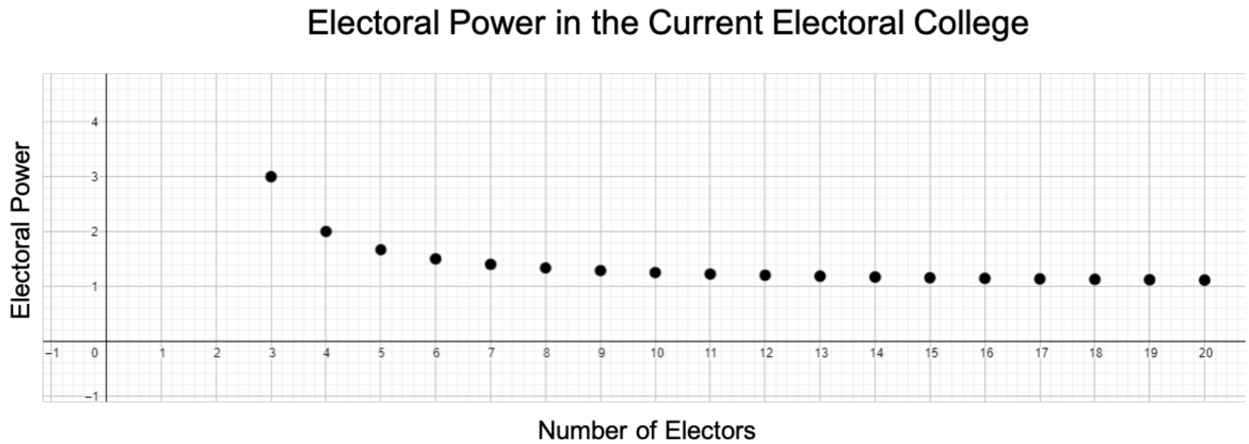
First, we fill in the number of representatives, senators and electoral votes for each state that we found earlier. Then we will calculate the number of electoral votes per 40,000 people.

To do that, divide the number of electors by the number of representatives, since they each represent approximately 40,000 people. The results are shown in the table below. Note this is approximate because we cut off the decimal remainders when calculating the number of representatives.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes Per 40,000 people
Small	154,600	3	2	5	$5 \div 3 \approx 1.67$
Medium	262,340	6	2	8	$8 \div 6 \approx 1.33$
Large	581,135	14	2	16	$16 \div 14 \approx 1.14$
Huge	1,362,070	21	2	23	$23 \div 21 \approx 1.10$

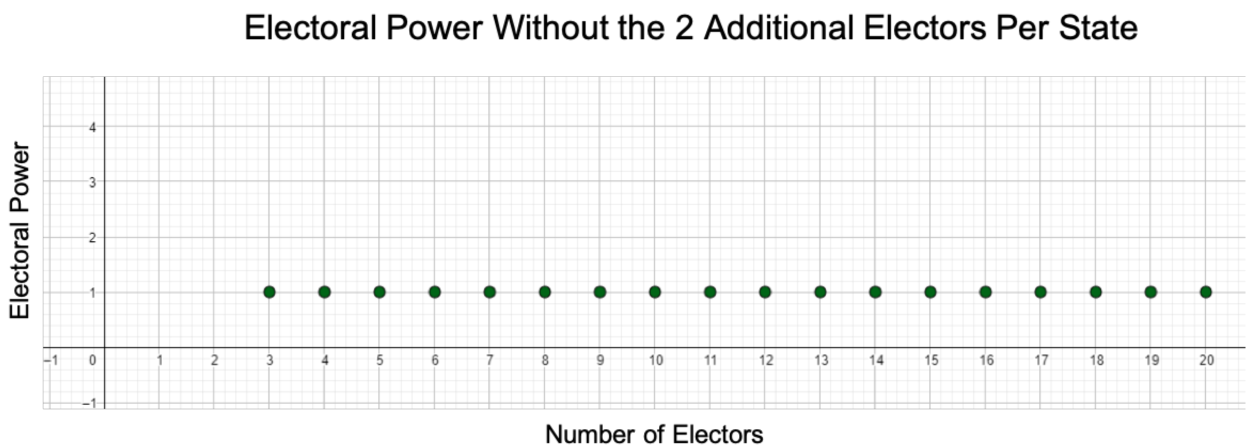
Are you surprised to see that the smallest states have the highest electoral power? That is part of the design of the electoral college. The founders from small states were concerned that their votes would not matter, so every state gets 2 senators added to the number of representatives. For a state with few representatives, those 2 extra electoral votes make a big difference. For a large states, the extra 2 electoral votes don't make a very big difference.

You may have noticed a pattern when dividing to find the electoral power. Each time we divided the number of electoral votes by that number minus two. We can write that pattern as $\frac{n}{n-2}$. Using GeoGebra to graph that pattern this is what we see.



The minimum number of electors is three, since each state gets at least one representative in the apportionment process. The smallest states and Washington, D.C. have 3 electors and the greatest electoral power on the left side of the graph. The largest states like California, Florida and Texas have the least electoral power on the right side of the graph.

Let's say instead each state had equal electoral power by removing the 2 extra electoral votes or by using the popular vote. Then the graph would look like this.



Now that we have studied the popular vote, the Electoral College and electoral power, we leave it up to you to choose which method you think is best. □

5.3.4 Exercises

1. How is the U.S. President selected?

2. How many electors are there in the U.S. Electoral College?

3. How many U.S. Senators represent each state?

4. What is the length of a term of a U.S. Senator?

5. How many Representatives serve in the U.S. House of Representatives?

6. What is the length of a term of a U.S. Representative?

In each fictional country in problems 7-10, use the rules of the U.S. government to complete the table and determine the following:

a. The total number of electors in the state.

b. The number of electoral votes needed for a majority and win a presidential election.

7. In this country there is one representative for every 50,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Gandhi	450,000			
Mandela	150,000			
Gbowee	600,000			
Total				

8. In this country there is one representative for every 75,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Johnson	225,000			
Rivera	450,000			
Milk	750,000			
Total				

9. In this country there is one representative for every 40,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Tamez	280,000			
Teters	200,000			
Herrington	400,000			
Osawa	360,000			
Total				

10. In this country there is one representative for every 60,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Johnson	120,000			
Jackson	480,000			
Cox	720,000			
Browne	420,000			
Total				

In each fictional country in problems 11-14, use the rules of the U.S. government (assume that all of a state's electoral votes go to the candidate who received the majority of the votes in that state) to complete the table and determine the following:

a. The winner of the popular vote in the country and the percentage of votes they won.

b. The winner of the electoral college who becomes the president and the percentage of electoral votes they won.

11. In this country from problem 7, there is one representative for every 50,000 residents.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Gandhi	216,000	234,000		
Mandela	37,500	112,500		
Gbowee	489,450	110,550		
Total Votes				

12. In this country from problem 8, there is one representative for every 75,000 residents.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Johnson	115,650	109,350		
Rivera	235,800	214,200		
Milk	117,750	632,250		
Total				

13. In this country from problem 9, there is one representative for every 40,000 residents.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Tamez	95,480	184,250		
Teters	104,200	95,800		
Herrington	203,600	196,400		
Osawa	46,080	313,920		
Total Votes				

14. In this country from problem 10, there is one representative for every 60,000 residents.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Johnson	52,440	67,560		
Jackson	418,080	61,920		
Cox	319,680	400,320		
Browne	350,280	69,270		
Total Votes				

In each fictional country in problems 15-18, use the rules of the U.S. government to complete the table and determine the following:

a. The state that has the most electoral power

b. The state that has the least electoral power

15. In this country from problem 7, there is one representative for every 50,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 50,000 people
Gandhi	450,000				
Mandela	150,000				
Gbowee	600,000				

16. In this country from problem 8, there is one representative for every 75,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 50,000 people
Johnson	225,000				
Rivera	450,000				
Milk	750,000				

17. In this country from problem 9, there is one representative for every 40,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 50,000 people
Tamez	280,000				
Teters	200,000				
Herrington	400,000				
Osawa	360,000				

18. In this country from problem 10, there is one representative for every 60,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 50,000 people
Johnson	120,000				
Jackson	480,000				
Cox	720,000				
Browne	420,000				

19. For the country in problem 7, determine all the possible combinations of states that would win the electoral college. What is the minimum number of votes needed to win the electoral college?

20. For the country in problem 8, determine all the possible combinations of states that would win the electoral college. What is the minimum number of votes needed to win the electoral college?

21. For the country in problem 9, determine all the possible combinations of states that would win the electoral college. What is the minimum number of votes needed to win the electoral college?

22. For the country in problem 10, determine all the possible combinations of states that would win the electoral college. What is the minimum number of votes needed to win the electoral college?

5.4 Gerrymandering and How to Measure It

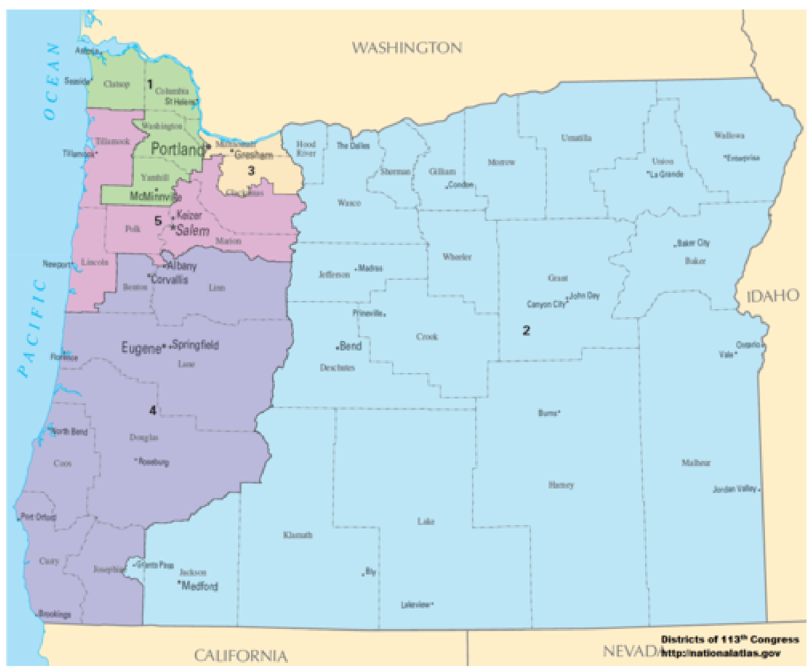
Objectives: Section 5.4

Students will be able to:

- Determine fair representation based on the proportion of each party in a state
- Explain what gerrymandering is and how packing and cracking work
- Calculate the efficiency gap for a given map
- Calculate the percentage of the population that each seat represents
- Determine the number of seats that the efficiency gap represents and whether a map is fair
- Draw district boundaries to form a gerrymandered map and a fair map

5.4.1 Forming State Legislative Districts

As we saw in Section 5.1 the U.S. House of Representatives has 435 seats that are re-apportioned among the states every 10 years after the census. For example, in Oregon we currently have 5 legislative districts as shown in this map[1]. After the 2020 census is finished, that number could change. It has been projected that Oregon may get a 6th representative due to population growth from 2010 to 2020.



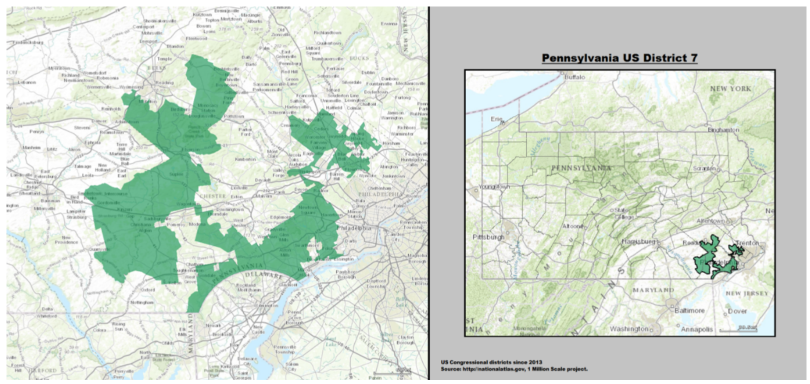
After a reapportionment, it is up to each state government to divide their population into equal districts which each elect their representative to the U.S. House. Each representative should represent approximately the same number of people, but their areas may not be the same. Due to different population densities in urban and rural areas, you can see in the map of Oregon that the districts are not geographically equal.

Currently in most states, the state legislature draws the district boundaries. That means the party in power is in charge of drawing the new districts and they could redraw the lines to help their own party. This is called gerrymandering. ***Gerrymandering***, pronounced, "Jerrymandering," is when districts are drawn to the political advantage of those drawing the boundaries.

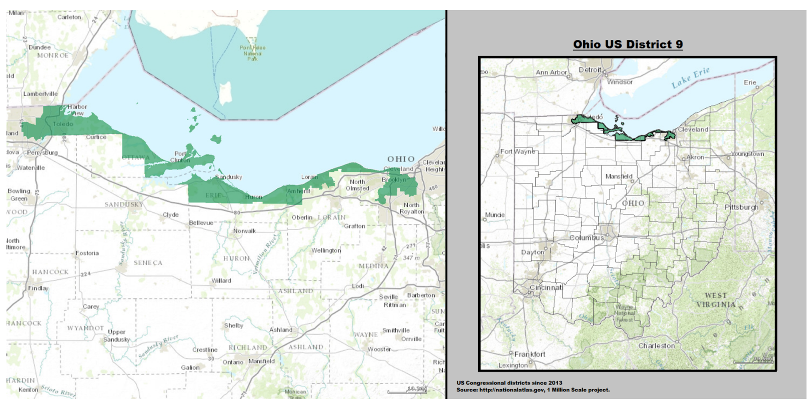
[1] "Oregon congressional districts" by National Atlas of the United States of America is in the Public Domain

Gerrymandering got its name from Elbridge Gerry, the governor of Massachusetts who signed a bill in 1812 that created the unusual district pictured here that looked like a salamander[2]. It gave then Democrat-Republicans a bigger state Senate majority, even though the Federalist party got more votes statewide.

Here are some current well-known gerrymanders. This one is called Pennsylvania's "Goofy Kicking Donald Duck"[3]



This one is called Ohio's "Lake Erie Monster"[4].



Gerrymandering has occurred by both parties. Gerrymandering by race is illegal due to the Voting Rights Act of 1965, so people of color cannot be spread out to dilute their vote. Groups with similar interests may want to be put together to form voting blocks, but that is controversial. Gerrymandering by political party has been the subject of many lawsuits and Supreme Court cases. One of the challenges is how to measure gerrymandering to prove that it has occurred. The measurement must be easy to explain and understand in court.

Several ways to measure gerrymandering have been proposed. Some of the first methods used geometry to measure how compact a district is, rather than the sprawling salamander. But it has been shown that being compact does not equate with fairness. More recent methods use computer simulations to determine whether

[2]"The Gerry-Mander" by Elkanah Tisdale (1771-1835), Originally published in the Boston Centinel, 1812 is in the Public Domain

[3]"Pennsylvania US Congressional District 7 (since 2013).tif" by Wikimedia Commons contributors, Wikimedia Commons, the free media repository. is in the Public Domain

[4]"Ohio US Congressional District 9 (since 2013).tif" by Wikimedia Commons contributors, Wikimedia Commons, the free media repository. is in the Public Domain

a certain map is an extreme outlier compared to other possible maps, or measure something called partisan bias[5].

In this book we will focus on the efficiency gap, proposed in 2015 by Nicholas Stephanopoulos and Eric McGhee. The ***efficiency gap*** is a measure of the advantage one party has over the other party due to the partisanship of the voters in each district.

Later in this section we will learn how to calculate and interpret the efficiency gap. First, let's look at how we can tell whether a map is fair or not.

5.4.2 Proportionality and Fairness

For a map to be fair, we would expect the portion of seats that a political party wins in an election to be about the same as the portion of that party in the state. This is called ***proportionality***. For example, if a state is 50:50 Democrat to Republican, they should each get about half of the seats. There is nothing in the constitution that guarantees proportionality, but this is a basic measure of fairness. Let's look at a couple of examples using fractions and decimals.

Example 5.4.1 If a state with 6 seats is $\frac{1}{3}$ Republican and $\frac{2}{3}$ Democrat, how many seats would each party have if the representation was proportional?

We will multiply the total number of seats by fraction that each party represents.

Republicans: $6 \cdot \frac{1}{3} = \frac{6}{1} \cdot \frac{1}{3} = \frac{6}{3} = 2$ seats.

Democrats: $6 \cdot \frac{2}{3} = \frac{6}{1} \cdot \frac{2}{3} = \frac{12}{3} = 4$ seats.

In this case the Republicans should have 2 seats and the Democrats should have 4 seats. The numbers won't usually work out so nicely so let's look at another example. □

Example 5.4.2 If a state with 8 seats is 55% Republican and 45% Democrat, how many seats would each party have if the representation was proportional?

Solution. We will multiply the total number of seats by the decimal portion that each party represents.

Republicans: $8(0.55) = 4.4 \approx 4$ seats.

Democrats: $8(0.045) = 3.6 \approx 4$ seats.

By rounding, the most fair combination would be Democrats with 4 seats and Republicans with 4 seats. This is not completely proportional, though, because we can't assign fractions of seats. □

Now that we can tell whether a map is fair, let's see how gerrymandering works.

5.4.3 How to Gerrymander

There are two main ways to gerrymander a map, by packing and cracking. ***Packing*** is when all the people of one political party are packed into a district. Since a party only needs a plurality or majority to win, all those extra votes would be surplus. ***Cracking*** dilutes a party's votes by spreading out the voters so they can't win as many districts.

[5] https://projects.fivethirtyeight.com/partisan-gerrymandering-north-carolina/

Here is a small sample map to see how this works. In this fictitious state, we have 35 people and 5 districts. So we will need 7 people in each district. You can see how they are spread out in the map, with Republicans denoted by R's and Democrats marked with D's. To keep it simple, we are only looking at two parties and they will need a majority to win an election. We are also assuming that all voters will vote and that they vote according to their party which may not always be the case.

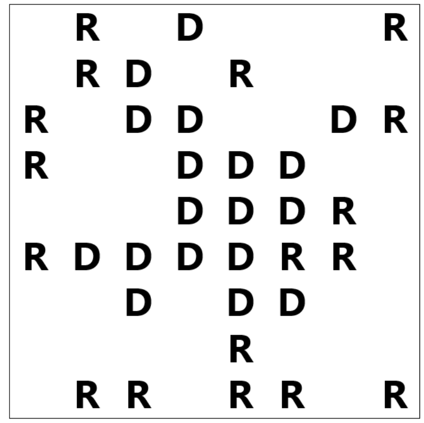


Figure 5.4.3 A simplified state map

Before we draw a fair map, let's try gerrymandering it for Democrats and then for the Republicans.

Example 5.4.4

We will first try to advantage the Democrats. There are 7 people in each district, so a party needs 4 votes to win the election. We will draw the lines to put 4 Democrats in each district along with 3 Republicans to spread them out by cracking. Here is our map. We were able to get 4 Democrats and 3 Republicans in districts 1, 2, 3 and 5. The Republicans will win in district 4 because it will be 5-2, but if the 5th Republican had been in another district, the map would have been more fair. As it is drawn, the population is 17:18 but the seats are 1:4, which does not seem fair.

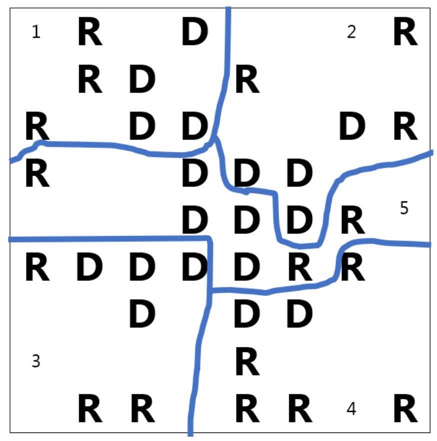


Figure 5.4.5 The map gerrymandered in favor of Democrats

□

5.4.4 The Efficiency Gap

To try to measure how unfair the map is, let's learn how to calculate the efficiency gap. First, we will tally the votes in the table below, and note how the election results would turn out by using bold numbers for the party with more voters in each district. For example, there are 4 D's and 3 R's in district 1, so the Democrats would theoretically win that district.

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	**4**	3		
2	**4**	3		
3	**4**	3		
4	2	**5**		
5	**4**	3		
Total	18	17		

Election Results:

- Democrats win 4 seats

- Republicans win 1 seat

Next, we will calculate how many votes would be extra over the amount needed to win the election. These are the ***surplus votes***. In this example, with 7 voters, the majority of a district is 4 votes, so we will subtract 4 from each winning side. From each losing side, all of the votes are considered surplus because they did not go toward electing a candidate of their party. Then we add all the surplus votes for each party.

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	***4***	3	$4-4=0$	3
2	***4***	3	$4-4=0$	3
3	***4***	3	$4-4=0$	3
4	2	***5***	2	$5-4=1$
5	***4***	3	$4-4=0$	3
Total	18	17	2	13

The difference between the surplus votes as a percentage of the population is the efficiency gap. As a formula it is written:

$$\frac{\text{Party A Surplus Votes} - \text{Party B Surplus Votes}}{\text{Total Votes}}$$

To get a positive result, we put the larger number first and we have

$\frac{13-2}{35} = \frac{11}{35} \approx 0.314$ or 31.4%.

The efficiency gap is 31.4%. To understand what this percentage means, we compare it with the percentage that each district represents in the state. Since our state has 5 districts, each one is 20% of the population. Here's how we got that by dividing or as a fraction of the state.

$100\% \div 5$ districts $= 20\%$, or, $\frac{1}{5} = 0.20$ or 20%

In this case, the efficiency gap is worth more than one full seat, which suggests that the Democratic party has at least one extra seat than they would with proportional representation.

Example 5.4.6

For another example, we will take the same map and gerrymander it for the Republicans. This time we will pack as many Democrats as we can into the middle, and then use cracking to form the rest of the districts, trying to get 4 R's in as many as we can. Here is the map.

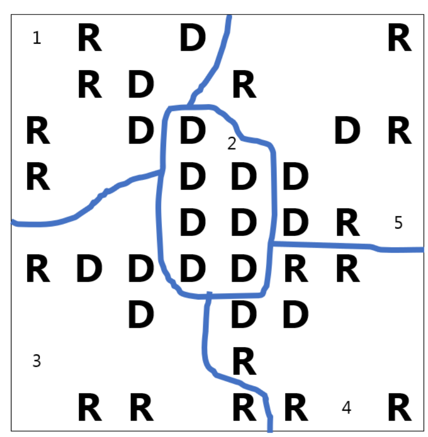


Figure 5.4.7 The map gerrymandered in favor of Republicans

Now, let's fill in our table:

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	3	4	3	$4-4=0$
2	7	0	$7-4=3$	0
3	3	4	3	$4-4=0$
4	2	5	2	$5-4=1$
5	3	4	3	$4-4=0$
Total	18	17	14	1

Election Results:

- Democrats win 1 seat
- Republicans win 4 seats

To calculate our efficiency gap,

$\frac{\text{Party A Surplus Votes} - \text{Party B Surplus Votes}}{\text{Total Votes}} = \frac{14-1}{35} = \frac{13}{35} \approx 0.371$ or 37.1%.

This shows the opposite unfairness where the population is 17:18 but the seats are 4:1. We can calculate the number of seats that the efficiency gap is worth by dividing 37.1% by 20%.

$37.1\% \div 20\% = 1.855$

The efficiency gap is worth approximately 1.9 seats.

Stephanopoulos and McGhee gave a guideline of around 8% or less for the efficiency gap and this one is 37.1%. The Republicans have at least one extra seat than they would if the seats were proportional to the population. □

Example 5.4.8

Now we will try to create a fair map and see what the efficiency gap is. There is an odd number of districts and the population is 17:18 so it might be challenging to make it fair.
With this map, we tried to have two districts that favored Democrats, two that favored Republicans, and one that might go either way. Since there is one extra Democrat, the population is 17:18 and the seats are 2:3

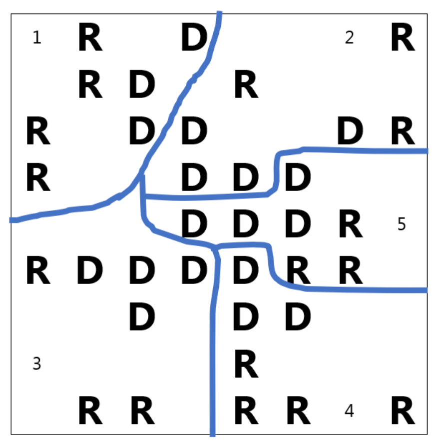


Figure 5.4.9 A more fair map

Let's see how the efficiency gap comes out in this situation:

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	3	4	3	$4-4=0$
2	4	3	$4-4=0$	3
3	4	3	$4-4=0$	3
4	3	4	3	$4-4=0$
5	4	3	$4-4=0$	3
Total	18	17	6	9

Election Results:

- Democrats win 3 seats
- Republicans win 2 seats

The efficiency gap is

$\frac{\text{Party A Surplus Votes} - \text{Party B Surplus Votes}}{\text{Total Votes}} = \frac{9-6}{35} = \frac{3}{35} \approx 0.0857$ or 8.57%.

This gap is much smaller than in our gerrymandered examples and near the 8% guideline. Due to the odd number of seats and even distribution of the parties, this is as small as we can get the efficiency gap in this situation. It is much less than 20%, which is one seat.

$8.57\% \div 20\% = 0.4285$

The efficiency gap is worth less than half a seat or about 0.43 seats. □

There are still many court cases in progress alleging gerrymandered maps so it seems like a new solution is needed. Some alternatives are appointing independent commissions to draw the lines, or changing the system altogether with proportional representation [6]. The statisticians at FiveThirtyEight used a web application to draw new lines in all of the states according to six different measures. You can read about it in this article [7].

5.4.5 Exercises

1. When does redistricting of state districts happen?

2. Who determines where the district lines are drawn?

3. What are the two rules for drawing district lines?

4. What are the two ways to gerrymander?

5. If a state with 7 seats is 62% Republican and 38% Democrat, how many seats would each party have if the representation was proportional?

6. If a state with 12 seats is 43% Republican and 57% Democrat, how many seats would each party have if the representation was proportional?

7. If a state with 3 seats is 4/5 Republican and 1/5 Democrat, how many seats would each party have if the representation was proportional?

8. If a state with 9 seats is 3/8 Republican and 5/8 Democrat, how many seats would each party have if the representation was proportional?

9. If a state with 11 seats is 52% Republican, 40% Democrat and 8% Green Party, how many seats would each party have if the representation was proportional?

10. If a state with 18 seats is 31% Republican, 58% Democrat and 11% Progressive Party, how many seats would each party have if the representation was proportional?

For each map in problems 11-20, complete the following:

a. How many votes are needed for a majority?

b. How many seats are won by each party?

c. Calculate the efficiency gap.

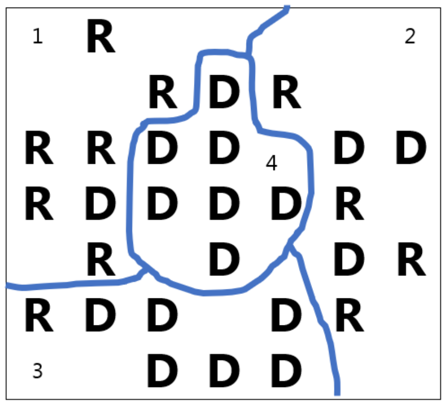
d. Calculate the percentage of the state that each district represents.

e. Calculate how many district seats the efficiency gap is worth.

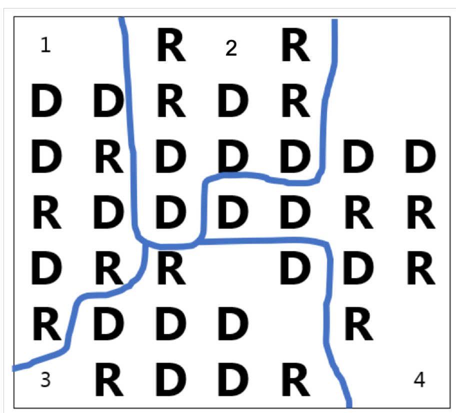
[6] https://www.fairvote.org/fair_representation#what_is_fair_voting
[7] https://fivethirtyeight.com/features/hating-gerrymandering-is-easy-fixing-it-is-harder/

f. Explain whether you think the map is fair and why or why not.

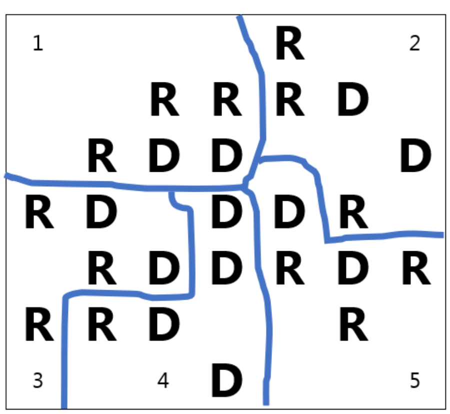
11. This state has 4 districts with 7 people in each district.



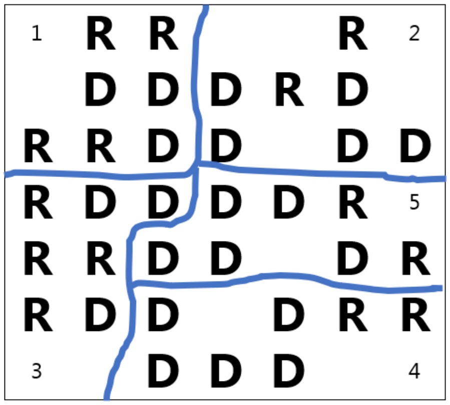
12. This state has 4 districts with 9 people in each.



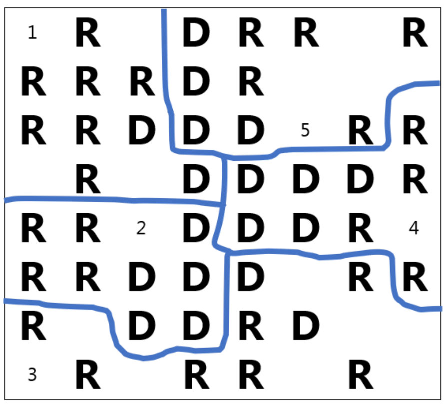
13. This state has 5 districts with 5 people in each.



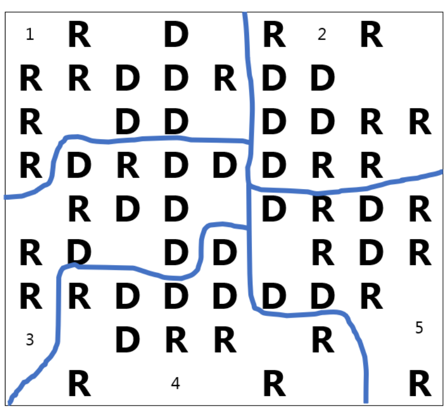
14. This state has 5 districts with 7 people in each.



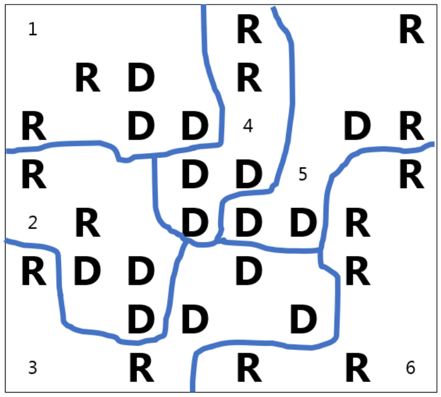
15. This state has 5 districts with 9 people in each.



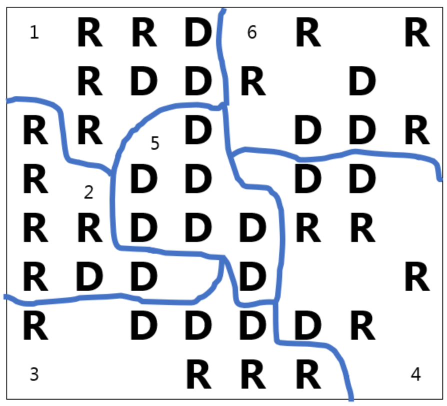
16. This state has 5 districts with 11 people in each.



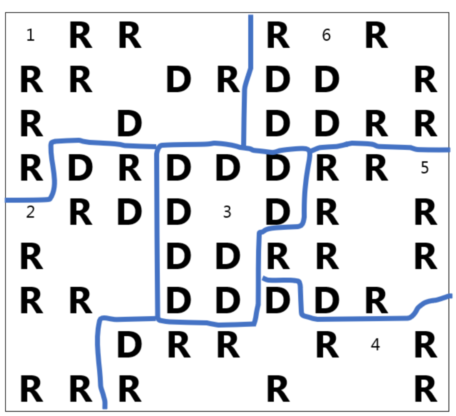
17. This state has 6 districts with 5 people in each.



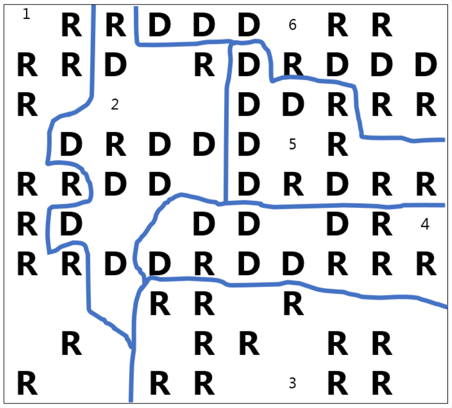
18. This state has 6 districts with 7 people in each.



19. This state has 6 districts with 9 people in each.



20. This state has 6 districts with 11 people in each.

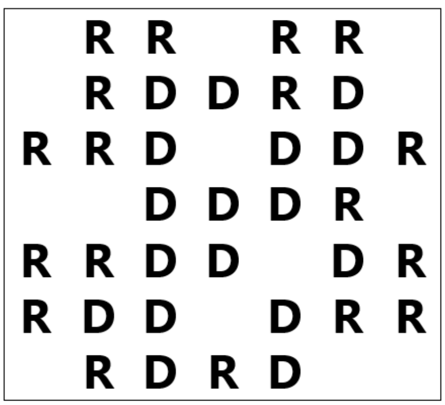


For each map in problems 21-24, draw your own districts to find

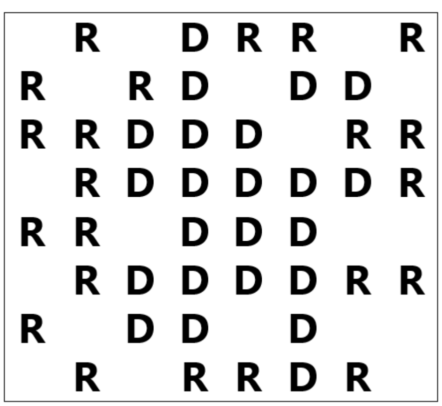
a. The most seats you can win for the Democrats.

b. The most seats you can win for the Republicans.

21. This state has 5 districts with 7 people in each.



22. This state has 5 districts with 9 people in each.



23. This state has 6 districts with 7 people in each.

	R	R			R		R
	R		D	R			
R	R	R	D	D	D	D	R
R		D	D		D	D	
R	R	D	D	D	R	R	
	D	D		D	R		R
R		D	D	D	D	R	
		R	R	R	R		

24. This state has 6 districts with 9 people in each.

	R	R			R		R	
	R		D	D	D	D		R
R		D	D	D	D	D		R
R	D	D	D	D	D	D	R	
	R	D	D		D	D	D	D
R		D	D	D		R		R
R	R		D	D	D	D	R	
	D	D	D	R		R		R
	R	R			R			

5.5 Chapter 5 Review

Review Exercises

In exercises 1-4, determine the apportionment using

a. Hamilton's Method

b. Jefferson's Method

c. Webster's Method

d. Huntington-Hill Method

1. A small country consists of four states, whose populations are listed below. If the legislature has 78 seats, apportion the seats.

A: 96,400	B: 162,700	C: 119,900	D: 384,900

2. Reapportion the previous problem with 90 seats.

3. A small country consists of five states, whose populations are listed below. If the legislature as 100 seats, apportion the seats.

A: 584,000	B: 226,600	C: 88,500	D: 257,300	E: 104,300

4. Reapportion the previous problem with 125 seats.

In exercises 5-8, complete the following:

a. How many voters voted in this election?

b. How many votes are needed for a majority?

c. Find the winner under the plurality method.

d. Find the winner under the Instant Runoff Voting method.

e. Find the winner under the Borda Count Method.

f. Find the winner under Copeland's method.

5. A Portland Community College Board member race has four candidates: E, F, G, H. The votes are:

Number of voters	12	16	17	15	34	13	19	8
1st choice	G	H	E	E	F	G	H	G
2nd choice	E	F	F	H	G	H	G	F
3rd choice	F	G	G	F	H	E	F	E
4th choice	H	E	H	G	E	F	E	H

6. A Forest Grove School Board position has four candidates: I, J, K, L. The votes are:

Number of voters	15	13	25	16	18	10	7	11	2
1st choice	K	I	J	L	K	L	I	I	L
2nd choice	J	L	L	I	I	J	K	J	K
3rd choice	L	J	I	K	J	I	J	K	J
4th choice	I	K	K	J	L	K	L	L	I

7. A Multnomah County Commissioner's race has five candidates: M, N, O, P, Q. The votes are:

Number of voters	31	18	35	37	33	12
1st choice	M	Q	O	N	P	Q
2nd choice	P	O	Q	P	M	N
3rd choice	O	M	P	O	N	M
4th choice	N	P	N	M	Q	O
5th choice	Q	N	M	Q	O	P

8. The Oregon State Governor's race has five candidates: R, S, T, U, V. The votes are:

Number of voters	22	45	20	47	43	18	26
1st choice	R	S	R	U	T	V	V
2nd choice	T	V	S	T	U	S	T
3rd choice	S	T	V	S	V	U	S
4th choice	U	R	U	V	R	R	U
5th choice	V	U	T	R	S	T	R

In each fictional country in problems 9-10, use the rules of the U.S. government to complete the table and determine the following:

a. The total number of electors in the state.

b. The number of electoral votes needed for a majority and win a presidential election.

9. In this country there is one representative for every 55,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Fonville	825,000			
Gurley	550,000			
Nevarez	275,000			
Total				

10. In this country there is one representative for every 60,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Arbery	720,000			
Monterrosa	360,000			
Bland	240,000			
Davis	480,000			
Total				

In each fictional country in problems 11-12, use the rules of the U.S. government (assume that all of a state's electoral votes go to the candidate who received the majority of the votes in that state) to complete the table and determine the following:

a. The winner of the popular vote in the country and the percentage of votes they won.

b. The winner of the electoral college who becomes the president and the percentage of electoral votes they won.

11. In this country from problem 9, there is one representative for every 55,000 residents.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Fonville	684,750	140,250		
Gurley	257,400	292,600		
Nevarez	132,275	142,725		
Total Votes				

12. In this country from problem 10, there is one representative for every 60,000 residents.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Arbery	372,240	347,760		
Monterrosa	38,880	321,120		
Bland	134,640	105,360		
Davis	104,160	375,840		
Total				

In each fictional country in problems 13-14, use the rules of the U.S. government to complete the table and determine the following:

a. The state that has the most electoral power

b. The state that has the least electoral power

13. In this country from problem 9, there is one representative for every 55,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 55,000 people
Fonville	825,000				
Gurley	550,000				
Nevarez	275,000				

14. In this country from problem 10, there is one representative for every 60,000 residents.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 60,000 people
Arbery	720,000				
Monterrosa	360,000				
Bland	240,000				
Davis	480,000				

For each map in problems 15-16, complete the following:

a. How many votes are needed for a majority?

b. How many seats are won by each party?

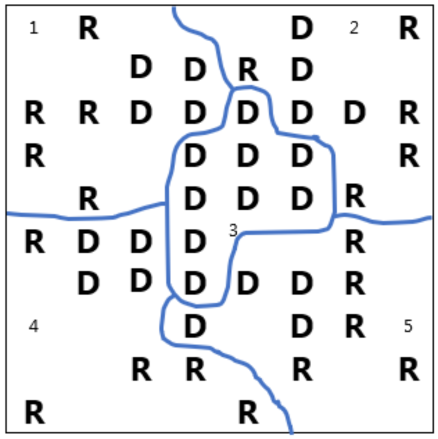
c. Calculate the efficiency gap.

d. Calculate the percentage of the state that each district represents.

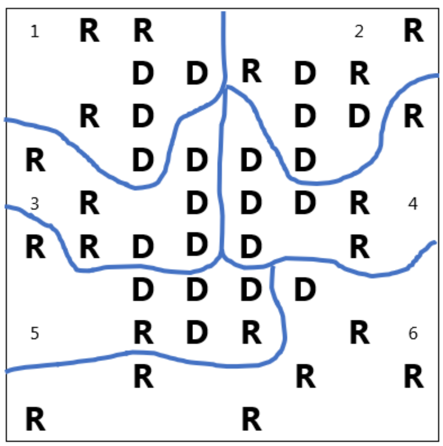
e. Calculate how many district seats the efficiency gap is worth.

f. Explain whether you think the map is fair and why or why not.

15. This state has 5 districts with 9 people in each district.



16. This state has 6 districts with 7 people in each.



"

5.6 Federal Budget, Deficit and National Debt

Objectives: Section 5.6

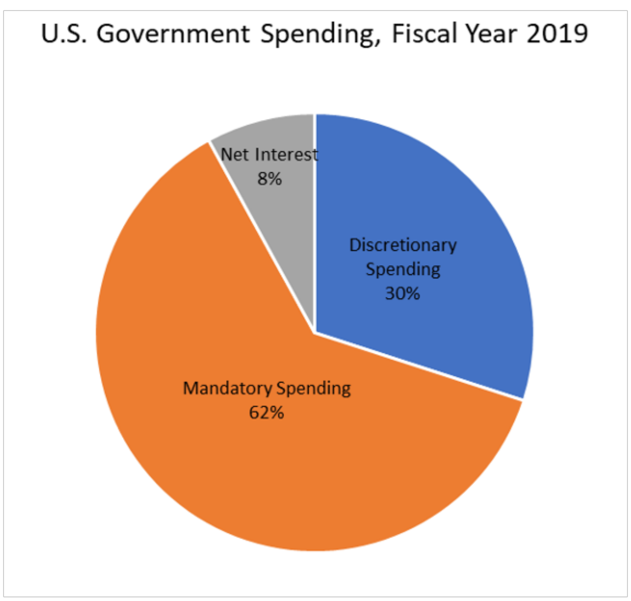
Students will be able to:

- Explain the federal budget process
- Explain the difference between the federal deficit and national debt
- Convert large numbers from expanded form to a decimal of millions, billions, trillions, etc. and back to expanded form
- Convert large numbers between decimals of millions to billions or trillions, etc.
- Calculate the debt to GDP ratio for a country or state
- Calculate the debt per person for a country or state
- Read and use pie charts to do calculations

Another important part of democracy is how to fund the government. We studied how to calculate federal income taxes in Section 2.5. In this section we will study where that money goes and what happens when the government spends more than it earns in taxes.

5.6.1 Federal Income and Spending

Federal income comes from our individual income taxes and business income taxes. Income is also borrowed by selling savings bonds, notes and Treasury bills (United States Government, 2019). The government has to pay interest on the national debt. After that, there are two types of spending, mandatory and discretionary. ***Mandatory spending*** includes Social Security payments, Medicare and other items required by law. ***Discretionary spending*** is the amount that Congress budgets annually to fund programs and agencies. Here is a pie chart showing the percentage of spending for each type in 2019. The data comes from the Congressional Budget Office (Congressional Budget Office, 2020).



5.6.2 The Federal Budget Process

The ***Federal Budget*** is like a home budget, with income and expenses. The U.S. Government's fiscal year goes from October 1st of one year to September 30th of the next year. Work on the budget begins about a year and a half before the budget is finalized.

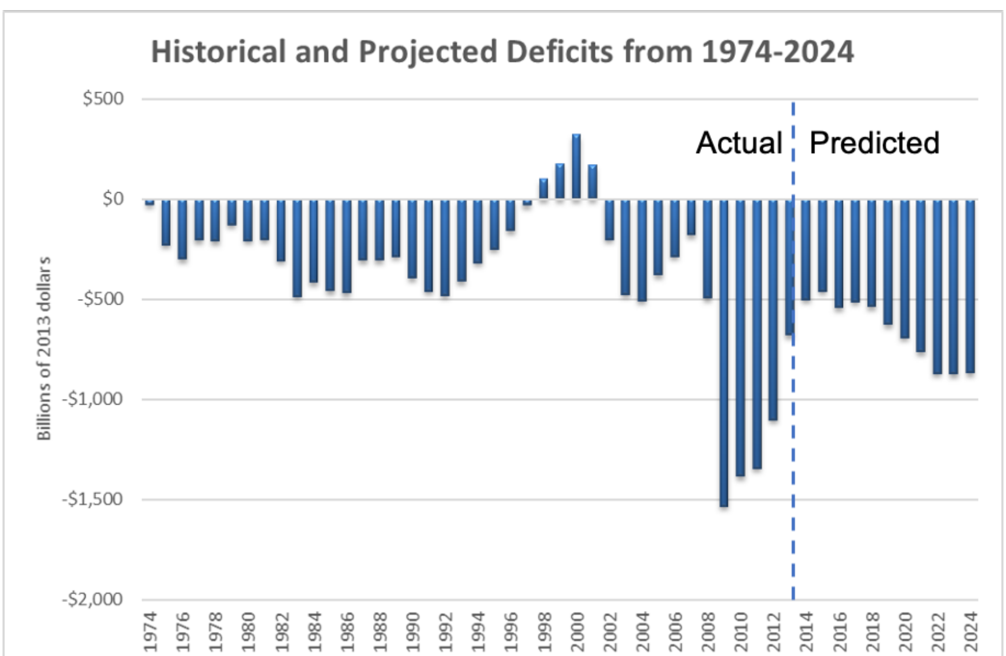
First, departments and agencies submit their proposals to the White House. Then the President submits their budget plan to Congress. Each chamber of Congress analyzes the proposal and makes their own budget resolution. Then a conference committee of House and Senate members resolves the differences between the two plans and makes a final version that each chamber votes on.

After the budget is passed, the Senate and House Appropriations committees distribute the discretionary part of the funding among 12 subcommittees that oversee different groups of agencies. As before, conference committees meet to merge the two versions of the appropriations bills. All 12 bills must be passed by both the Senate and the House and then signed by the President to enact the new budget.

If any appropriations bill is not signed by September 30th, there is no budget for the new fiscal year. In this case, Congress must pass a ***continuing resolution*** to temporarily fund the government. If they do not, or if the president does not sign it, the governement will shut down. The last ***government shutdown*** went from December 22, 2018 to January 25, 2019. This was the longest shudown in U.S. history and left over 800,000 federal workers either working without pay or being furloughed at home (Robert, 2020).

5.6.3 The Federal Surplus or Deficit

If the government spends less than it collects in income, there will be a budget ***surplus*** for that year. If it spends more than it collects, there will be a ***deficit*** for that year. It is also possible to have a ***balanced budget*** where the spending equals the income. The federal deficit refers to the budget shortfall in a single time period, like a quarter or a year. For example, in the Fiscal Year 2018, the U.S. deficit was $779 billion. The graph below shows the federal deficit or surplus each year since 1974 using data from the Congressional Budget Office (Congressional Budget Office, 2014).



It is hard to find graphs with a vertical axis in dollars becase the value of the dollar changes over time. One dollar in 1974 could buy a lot more than it can today. In the graph above, each dollar amount is converted to the equivalent of 2013 dollars to account for that. The vertical scale is in billions of 2013 dollars. The largest deficits in 2009 to 2012 were from spending and corporate bailouts to recover from the Great Recession of 2008.

On the graph above, \$1 on the vertical scale represents one billion dollars or \$1,000,000,000. The highest deficits go down to about -\$1,500 billion, which we would write as -\$1,500,000,000,000. Here is a chart that shows large numbers and how many zeros they have.

Number	Name
\$1,000	One thousand
\$1,000,000	One million
\$1,000,000,000	One billion
\$1,000,000,000,000	One trillion
\$1,000,000,000,000,000	One quadrillion
\$1,000,000,000,000,000,000	One quintillion

Instead of writing all the zeros, we can abreviate large numbers using decimals. For example, 1,200,000 can be written as 1.2 million. We can write 900,000 as 0.9 million. We can also translate the other way and write 4.567 trillion as 4,567,000,000,000. Note there can be more than one way to write a number. We could write 400,000,000 as either 400 million or 0.4 billion. It is convenient that we put the decimal where the comma goes and vice versa.

Example 5.6.1 Write each number using a decimal abreviation.

a 4,873,000

b 1,500,000,000

c 500,000,000,000

d 8,300,000

Solution. Our method is to find the millions place which is the 7th digit from the right. If the number starts there, use millions. If the number is larger, count 3 more places for billions and 3 more places for trillions.

a 4,873,000 = 4.873 million

b 1,500,000,000 = 1.5 billion

c 500,000,000,000 = 500 billion or 0.5 trillion

d 8,300,000 = 8.3 million

□

Example 5.6.2 Write each number in expanded form.

a 5.7 million

b 9.22 trillion

c 100.2 billion

d 0.25 trillion

Solution. Our method is to put a comma where the decimal is and then add zeros to get to the right place value. If the decimal is less than one, we move down to the next lower grouping as shown in part d.

a 5.7 million = 5,700,000

b 9.22 trillion = 9,220,000,000,000

c 100.2 billion = 100,200,000,000

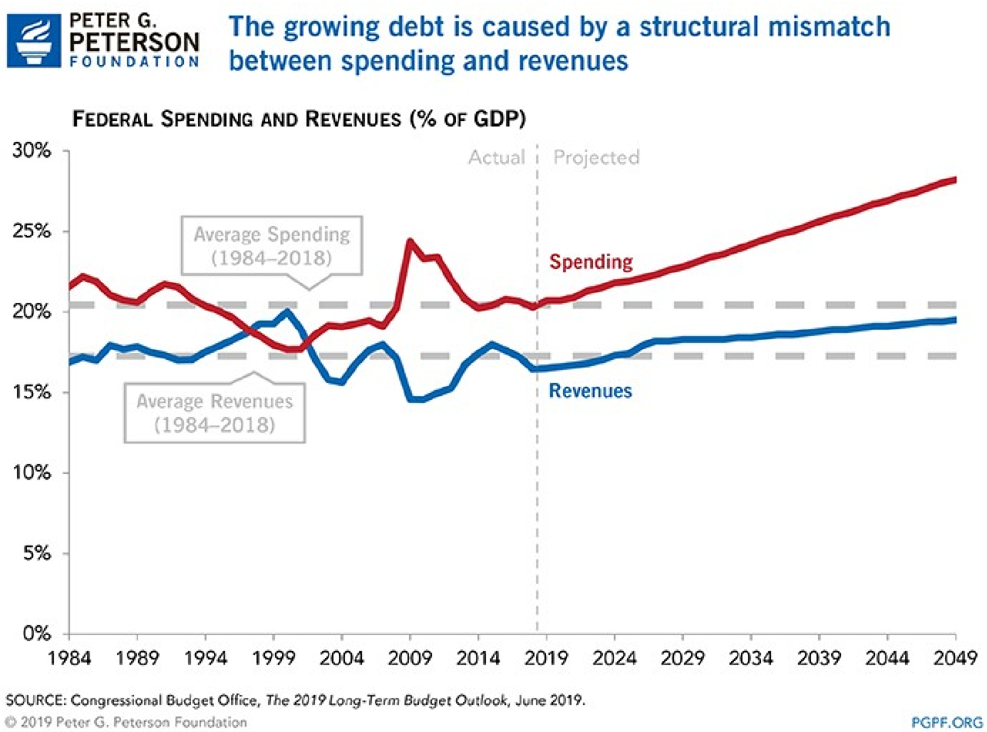
d 0.25 trillion = 250 billion = 250,000,000,000

□

5.6.4 Debt to GDP Ratio

A more common way to measure the deficit is as a percentage of the gross domestic product. The ***gross domestic product***, or ***GDP***, is the total value of all the finished goods and services produced within a county's borders in a specific period of time. The GDP is a measure of the size of an economy. The growth rate of the GDP is one measure of a nation's economic health (Bureau of Economic Analysis, 2019). The GDP of the United States in 2019 was $21.73 trillion but due to the COVID-19 pandemic, there was a drop in the second quarter of 2020 to $19.74 trillion. (Bureau of Economic Analysis, 2020).

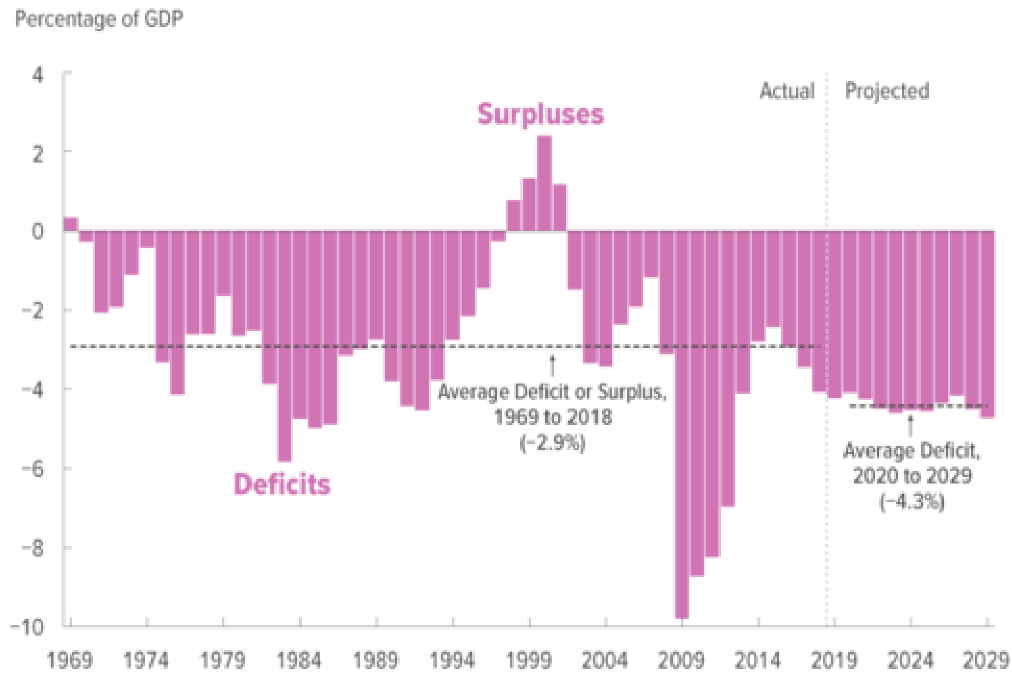
Now we can look at more graphs that are written in terms of the percent of GDP. Here is a graph of federal spending and revenues as a percentage of the GDP[1].



For each year, if we take the revenue and subtract the spending, we get the budget surplus or deficit. If the result is negative it is a deficit. And here is a graph of the deficits in terms of percentage of GDP[2].

[1] "Federal Revenue and Spending, 1985-2050" is copyrighted by Peter G. Peterson Foundation and used under the permissions granted for educational use.

[2] "Updated Budget Projections: 2019 to 2029" by Congressional Budget Office is in the Public Domain.

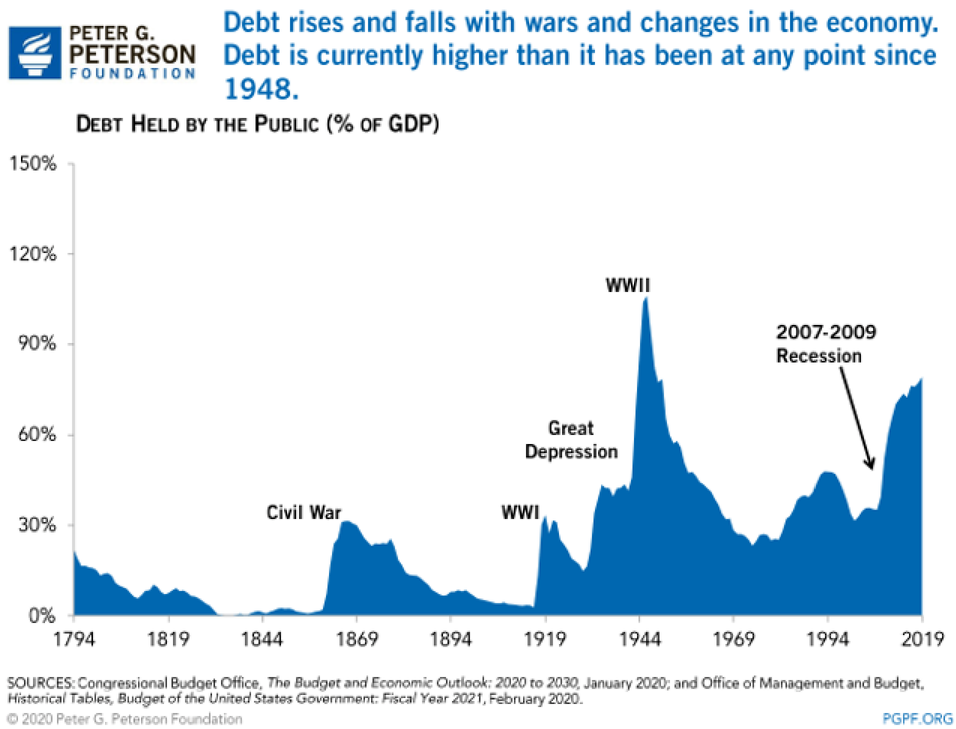


5.6.5 National Debt

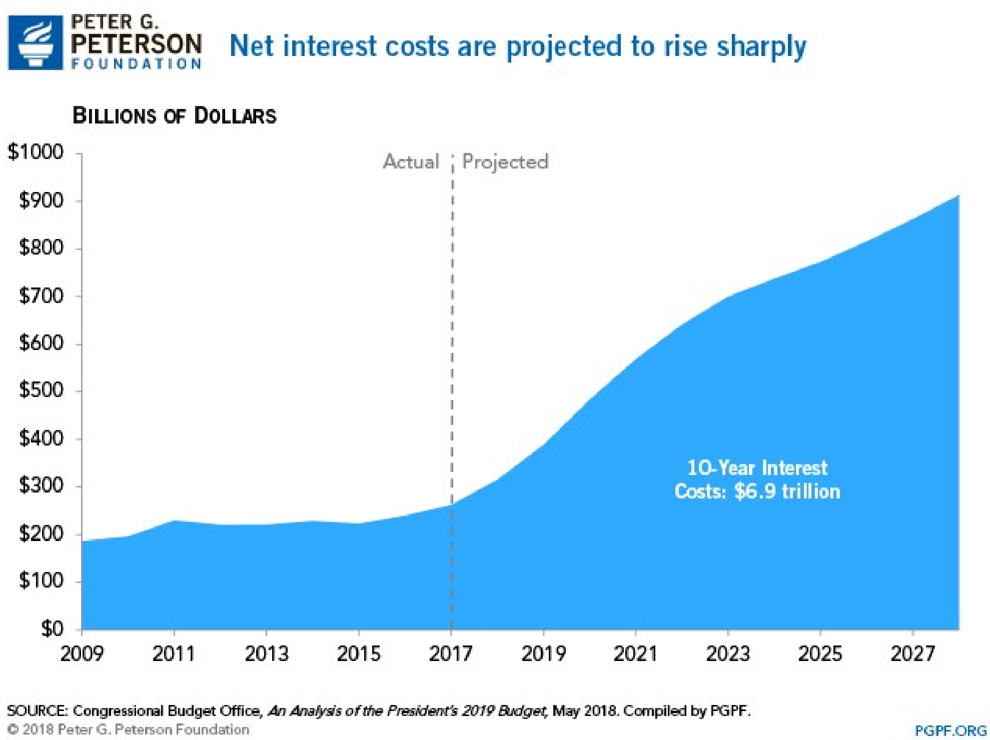
The words deficit and debt are easily confused because they have similar meanings. The ***deficit*** is the yearly shortfall, and the national ***debt*** is the total cumulative amount of debt held by the government.

As you can see from the previous graphs, most years had a budget deficit. If we add all of those up over time, this is what the national debt looks like[3].

[3]"Federal Debt: 1791-2019" is copyrighted by Peter G. Peterson Foundation and used under the permissions granted for educational use.



Just like we pay interest on our personal debt, the U.S. pays interest on the national debt. The interest on the debt is expected to keep increasing as shown in this graph[4].



[4] "Net Interest Cost" is copyrighted by Peter G. Peterson Foundation and used under the permissions granted for educational use.

5.6.6 National Debt Clocks

There are several websites that keep a running total of the national debt. One of them is USDebtClock.org. This one also shows the population, debt per citizen and debt per taxpayer. It shows the spending and deficit with counters that are constantly moving. At the bottom you can also see the debt to GDP ratio compared with previous ratios. The rest of the website shows statistics like the GDP, income tax revenue, median income, unemployment and much more. You can also see statistics for other states and countries.

Now let's do some calculations with these large numbers in the next example.

Example 5.6.3 Here are some approximate values for the U.S. from the Fiscal Year 2018: October 1, 2018 to September 30, 2019.

- Federal Budget (Spending): \$4.407 trillion
- Federal Revenue Estimate: \$3.422 trillion
- National Debt: \$21.803 trillion
- Interest on the National Debt for 2018: \$332.637 billion
- Gross Domestic Product: \$20.656 trillion
- Population: 329 million people

a Calculate the budget surplus or deficit for this fiscal year.

b Calculate the debt to GDP ratio as a percentage.

c How much debt does the U.S. have per person (per capita)?

d How much interest is due on the national debt per person?

Solution.

a To calculate the budget surplus or deficit, we subtract the spending from the revenue, and we get:

$$\begin{aligned}\text{Revenue } - \text{ Spending} &= \$3.422 \text{ trillion } - \$4.407 \text{ trillion}\\ &= -\$0.985 \text{ trillion or } -\$985 \text{ billion}\end{aligned}$$

The result is negative, so there is a deficit of \$985 billion. The word deficit indicates that the number is negative. If we wrote a deficit of -\$985 billion that would be incorrect because it would be a double negative.

b To calculate the debt to GDP ratio, we divide using the order of the wording. For example, the ratio of a to b would be $a : b$ or $a\ddot{o}b$. So we will take the total amount of national debt and divide it by the GDP:

$\frac{\text{national debt}}{\text{GDP}} = \frac{\$21.803 \text{ trillion}}{\$20.656 \text{ trillion}} = 1.06 \text{ or } 106\%.$

The debt to GDP ratio is 106%.

c "Per" is another key word for division, so to calculate the debt per capita or per person, we divide the national debt by the population.

$\frac{\text{national debt}}{\text{population}} = \frac{\$21.803 \text{ trillion}}{329 \text{ million people}}$

Notice that the debt is written in trillions and the population is in millions, so we can't divide these numbers yet. We need to put them into the same units. We can either write both of them in expanded form like this:

$\frac{\$21{,}803{,}000{,}000{,}000}{\$329{,}000{,}000} = \$66,271$ per person.

Or instead of writing out all the zeros, we can convert one of the numbers to match the other. In this situation it seems easier to convert \$21.803 trillion to 21,803,000 million and then divide.

$\frac{\$21.803 \text{ trillion}}{329 \text{ million people}} = \frac{\$21{,}803{,}000 \text{ million}}{329 \text{ million people}} = \$66,271$ per person.

d To find out how much interest is due on the national debt per person, we will divide the interest by the population.

$\frac{\text{interest}}{\text{population}} = \frac{\$332.637}{329 \text{ million people}}$

Again, the units do not match. This time we will convert the other way and change 329 million people into 0.329 billion people, and we have:

$= \frac{\$332.637 \text{ billion}}{0.329 \text{ billion}} = \$1,011.05$ per person.

Or you can always write out all the zeros like this and get the same answer.

$\frac{\$332{,}637{,}000{,}000}{329{,}000{,}000} = \$1,011.05$ per person.

□

Example 5.6.4 One advantage of the Debt to GDP ratio is you can compare different countries with economies of different sizes. Let's look at South Korea for comparison. Here are some approximate values, from 2020 (Commodity.com, 2020).

- Population: 50.617 million people
- National Debt: ₩754.835 trillion
- Gross Domestic Product: ₩1.731 quadrillion
- Interest Payments per year: ₩30.055 trillion

a Calculate the debt to GDP ratio as a percentage.

b Calculate the amount of debt per person.

c Calculate the amount of interest paid per year per person.

Solution.

a First, we divide the amount of debt by the GDP.

$\frac{\text{national debt}}{\text{GDP}} = \frac{₩754.835 \text{ trillion}}{₩1.71 \text{ quadrillion}}$

Since one of the numbers is in trillions and the other in quadrillions, we have to make them match before we can divide. We will change the GDP into trillions:

$\frac{₩754.835 \text{ trillion}}{₩1.71 \text{ quadrillion}} = 0.44$ or 44%.

b To find the amount of debt per person, we divde the debt by the number of people.

$\frac{\text{national debt}}{\text{population}} = \frac{₩754.835 \text{ trillion}}{50.617 \text{ million people}}$

Again the units don't match so we must either write all the zeros or make sure they are in the same units. In this case we will change the debt into millions but there are many ways you could do it.

$\frac{₩754.835 \text{ trillion}}{50.617 \text{ million people}} = ₩14,912,677.56$ per person.

c To find the amount of interest paid per year per person, we divde the amount of interest by the number of people.

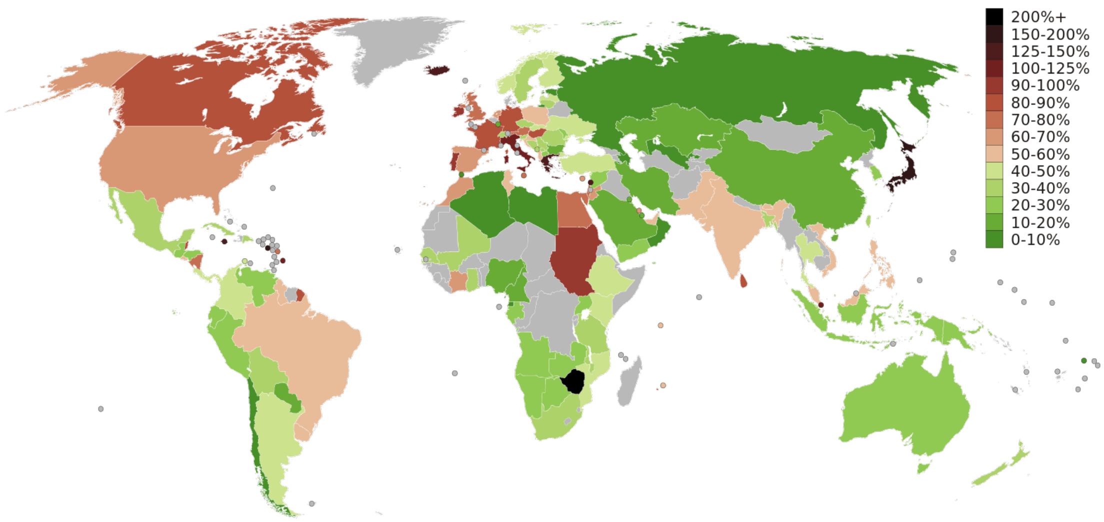
$\frac{\text{interest}}{\text{population}} = \frac{₩30.055 \text{ trillion}}{50.617 \text{ million people}}$

We will change the interest to millions and divide.

$\frac{₩30.055 \text{ trillion}}{50.617 \text{ million people}} = ₩593,772.84$ per person.

□

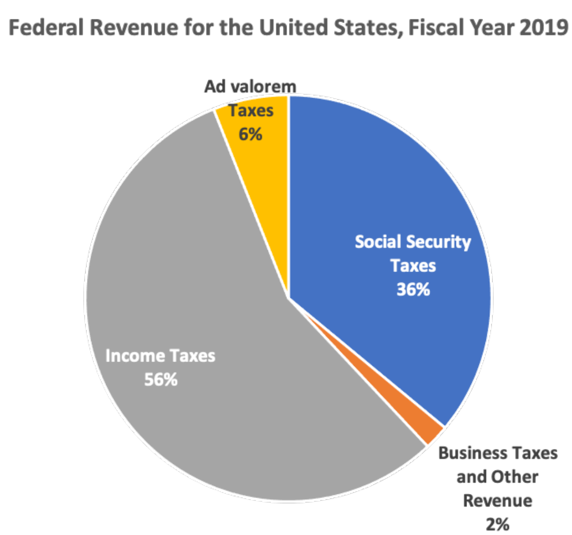
For more comparisons, here is a graph showing the debt to GDP ratios around the world in 2020[5].



5.6.7 Pie Charts and Percentages

We can also do calculations with large numbers that involve percentages. We will look at some pie charts related to government income and spending.

Example 5.6.5 In a previous example, we saw that the federal revenue estimate for 2019 was \$3.422 trillion. Use the pie chart to calculate the dollar value of each revenue source shown in the graph.



Solution. There are 4 segments in the pie chart, showing 4 different types of revenue. For each type, we will multiply the decimal form of the percentage by the total revenue:

Individual Income Taxes: (0.56)(\$3.422 trillion) = \$1.916 trillion

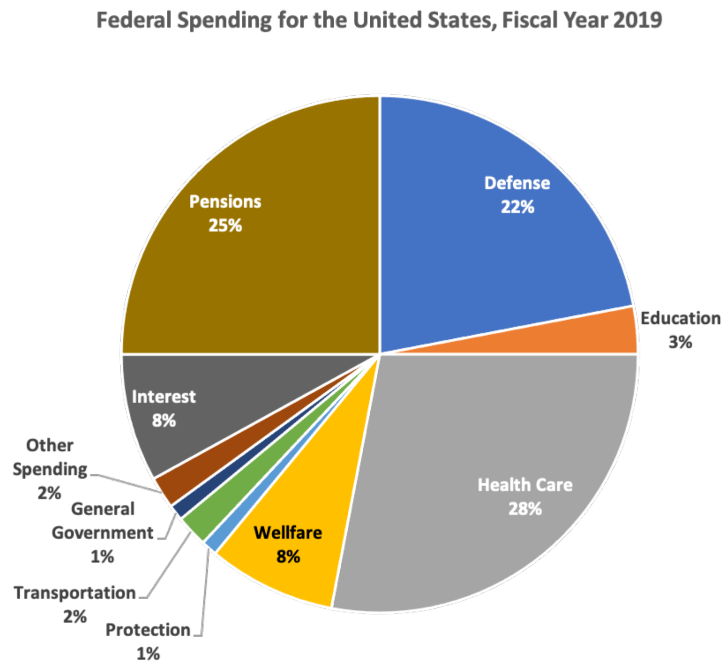
Social Security Taxes (Payroll Taxes): (0.36)(\$3.422 trillion) = \$1.232 trillion

[5]"Public Debt Percent GDP World Map" by Wikimedia Commons Contributors, Wikimedia Commons, the free media repository.

Corporate Taxes and Other: (.02)($3.422 trillion) = $0.0684 trillion or $68.4 billion

Ad Valorum (Excise Taxes and other): (.06)($3.422 trillion) = $205.3 billion □

Example 5.6.6 In a previsous example we also saw that the total amount of the federal budget in 2019 was $4.407 trillion. Use the pie chart of federal spending to calculate the dollar amounts of each of the following types of spending.



a Defense

b Social Security Payments

c Healthcare

d Education

Solution. We will take the decimal form of the percentage of each type of spending and multiply it by the total budget amount of $4.407 trillion.

a Defense: (0.22)($4.407 trillion) = $0.9695 trillion or $969.5 billion

b Social Security: (0.25)($4.407 trillion) = $1.102 trillion

c Healthcare: (0.28)($4.407 trillion) = $1.233 trillion

d Education: (0.03)($4.407 trillion) = $0.1322 trillion or $132.2 billion

□

In this chapter, we have looked at many important quantitative aspects of government: apportionment, voting methods, how the president is chosen, gerrymandering and the federal budget. Filling out the census, being informed and voting are extremely important for the U.S. democracy. Please make up your own mind and vote if you are eligible.

5.6.8 Exercises

1. Where does the U.S. federal income come from?

2. What are the two types of federal spending?

3. How many appropriations bills must be passed to approve the new federal budget?

4. When is the deadline for the new budget to be approved?

5. What is the difference between federal deficit and debt?

6. What is the Gross Domestic Product?

7. Write each number using a decimal abreviation.

a 4,300,000,000

b 12,567,000,000,000

c 500,000,000

d 6,040,000

8. Write each number using a decimal abreviation.

a 63,651,000,000,000

b 93,600,000

c 119,930,000,000

d 6,001,000,000

9. Write each number in expanded form.

a 5.7 million

b 9.22 trillion

c 100.2 billion

d 0.25 trillion

10. Write each number in expanded form.

a 0.52 quadrillion

b 1.49 billion

c 9.07 trillion

d 800 million

For each country in problems 11-16, find the following. Data from Commodity.com (Commodity.com, 2020).

a The debt to GDP ratio as a percentage.

b The amount of debt per person.

c The amount of interest paid per year per person.

11. In Columbia, the unit of currency is the Columbian peso, abbreviated as COP$ or C$.

- Population: 48.9 million people
- National Debt: C$ 270.978 trillion
- Gross Domestic Product: C$ 491.504 trillion
- Interest Payments per year: C$ 16.628 trillion

12. In Pakistan, the unit of currency is the Pakistani rupee or Rs.

- Population: 209.7 million people
- National Debt: Rs 24.255 trillion
- Gross Domestic Product: Rs 27.354 trillion
- Interest Payments per year: Rs 2.033 trillion

13. In Poland, the unit of currency is the złoty, or zł.

- Population: 38,492,299 people
- National Debt: 1.223 trillion zł
- Gross Domestic Product: 1.981 trillion zł
- Interest Payments per year: 52.500 million zł

14. In Australia, the unit of currency is the Australian dollar, AUD or A$.

- Population: 24.711 million people
- National Debt: A$ 645.316 billion
- Gross Domestic Product: A$ 1.927 trillion
- Interest Payments per year: A$ 19.73 billion

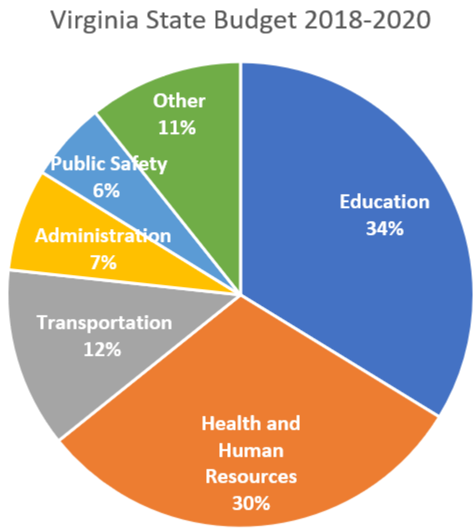
15. In South Africa, the unit of currency is the South African rand or R.

- Population: 54.5 million people
- National Debt: R. 3.736 trillion
- Gross Domestic Product: R. 6.186 trillion
- Interest Payments per year: R. 196.843 billion

16. In Malaysia, the unit of currency is the Malaysian ringgit or RM.

- Population: 54.5 million people
- National Debt: RM. 792.474 billion
- Gross Domestic Product: RM. 1.444 trillion
- Interest Payments per year: RM. 27.682 billion

17. The total state budget for the state of Virginia in 2018-2019 is approximately $63.9 billion and their population is about 8.5 million people. Use the pie chart (Virginia Department of Planning and Budget, 2018) to calculate how much is budgeted for:

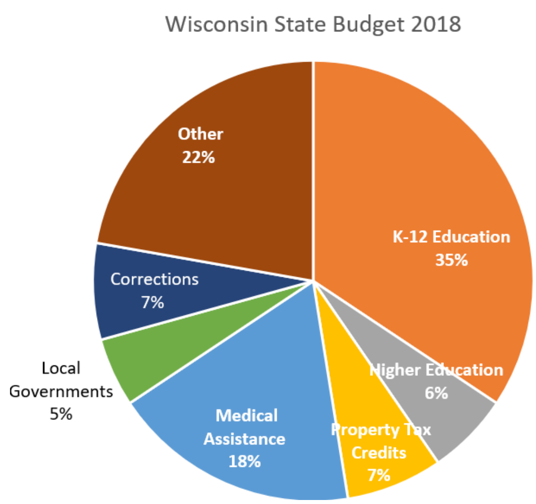


a Transportation in total

b Health and human resources per person

c Public safety in total

18. The Wisconsin state budget for 2018 was $8.9 billion and their population is about 5.8 million. Use the pie chart (Wisconsin Budget Project, 2020) to calculate how much was budgeted for:



a K-12 education in total

b Local governments per person

c Corrections in total

All Answers for Instructors Only

1 · Logic and Sets

1.1 · The Language and Rules of Logic

1.1.12 · Exercises

1.1.12.1.

a. Proposition

b. Not a proposition

c. Proposition

d. Proposition

1.1.12.2.

a. Not a proposition

b. Proposition

c. Not a proposition

d. Proposition

1.1.12.3.

a. I do not ride my bike to campus.

b. Portland is in Oregon.

1.1.12.4.

a. You should not see this movie.

b. Lashonda is not wearing blue.

1.1.12.5.

Answers will vary. Example: You can't not do your homework.

1.1.12.6.

Answers will vary. Example: They decided not to cancel the ban on pesticides.

1.1.12.7.

a. Exclusive

b. Inclusive

1.1.12.8.

a. Exclusive

b. Inclusive

1.1.12.9.

a. Inclusive

b. Exclusive

1.1.12.10.

a. Inclusive

b. Inclusive

1.1.12.11.

a. If it is sunny, then I will go swimming.

b. If it is Friday, then I will go see a movie.

1.1.12.12.

a. If it is raining, then I carry an umbrella.

b. If it is the weekend, then I am hanging out with friends.

1.1.12.13.

a. Elvis is not alive

b. Elvis is alive or Elvis is the King

c. Elvis is not alive and Elvis is the King

d. If Elvis is the King, then Elvis is not alive

1.1.12.14.

a. I do not own an umbrella.

b. It rains in Oregon and I do not own an umbrella

c. If it rains in Oregon, then I own an umbrella

d. If I do not own an umbrella, then it rains in Oregon

1.1.12.15.

a. B and A

b. If B, then A

c. If B, then A

d. If not B, then not A

1.1.12.16.

a. A and not B

b. Not A

c. If A, then B

d. If not B, then not A

1.1.12.17.

a. True

b. False

1.1.12.18.

a. False

b. False

1.1.12.19.

A	B	A and B
T	T	T - I live in Oregon and I go to PCC
T	F	F - I live in Oregon and I don't go to PCC
F	T	F - I don't live in Oregon and I go to PCC
F	F	F - I don't live in Oregon and I don't go to PCC

1.1.12.20.

A	B	A or B
T	T	T - I am a psychology major and I'm planning to transfer to PSU
T	F	T - I am a psychology major but I'm not planning to transfer to PSU
F	T	T - I am not a psychology major, but I'm planning to transfer to PSU
F	F	F - I am not a psychology major and I'm not planning to transfer to PSU

1.1.12.21.

A	B	Not B	A and not B
T	T	F	F
T	F	T	T
F	T	F	F
F	F	T	F

1.1.12.22.

A	B	Not A	Not A or B	Not (not A or B)
T	T	F	T	F
T	F	F	F	T
F	T	T	T	F
F	F	T	T	F

1.1.12.23.

A	B	C	A and B and C	Not (A and B and C)
T	T	T	T	F
T	T	F	F	T
T	F	T	F	T
T	F	F	F	T
F	T	T	F	T
F	T	F	F	T
F	F	T	F	T
F	F	F	F	T

1.1.12.24.

A	B	C	Not A	Not B	Not B and C	Not A or (Not B and C)
T	T	T	F	F	F	F
T	T	F	F	F	F	F
T	F	T	F	T	T	T
T	F	F	F	T	F	F
F	T	T	T	F	F	T
F	T	F	T	F	F	T
F	F	T	T	T	T	T
F	F	F	T	T	F	T

1.1.12.25.

A	B	C	A and B	Not (A and B)	Not (A and B) or C
T	T	T	T	F	T
T	T	F	T	F	F
T	F	T	F	T	T
T	F	F	F	T	T
F	T	T	F	T	T
F	T	F	F	T	T
F	F	T	F	T	T
F	F	F	F	T	T

1.1.12.26.

A	B	C	A or B	A or C	(A or B) and (A or C)
T	T	T	T	T	T
T	T	F	T	T	T
T	F	T	T	T	T
T	F	F	T	T	T
F	T	T	T	T	T
F	T	F	T	F	F
F	F	T	F	T	F
F	F	F	F	F	F

1.1.12.27.

A	B	C	A and B	If (A and B), then C
T	T	T	T	T
T	T	F	T	F
T	F	T	F	T
T	F	F	F	T
F	T	T	F	T
F	T	F	F	T
F	F	T	F	T
F	F	F	F	T

1.1.12.28.

A	B	C	A or B	Not C	If (A or B), then not C
T	T	T	T	F	F
T	T	F	T	T	T
T	F	T	T	F	F
T	F	F	T	T	T
F	T	T	T	F	F
F	T	F	T	T	T
F	F	T	F	F	T
F	F	F	F	T	T

1.1.12.29.

A	C	A and C	Not A	If (A and C), then not A
T	T	T	F	F
T	F	F	F	T
F	T	F	T	T
F	F	F	T	T

1.1.12.30.

A	B	C	B or C	A and B	If (B or C), then (A and B)
T	T	T	T	T	T
T	T	F	T	T	T
T	F	T	T	F	F
T	F	F	F	F	T
F	T	T	T	F	F
F	T	F	T	F	F
F	F	T	T	F	F
F	F	F	F	F	T

1.2 · Sets and Venn Diagrams

1.2.12 · Exercises

1.2.12.1.

{m, i, s, p}

1.2.12.2.

{January, February, March, April, May, June, July, August, September, October, November, December}

1.2.12.3.

The set containing multiples of 3 from 3 to 9

1.2.12.4.

The set containing vowels

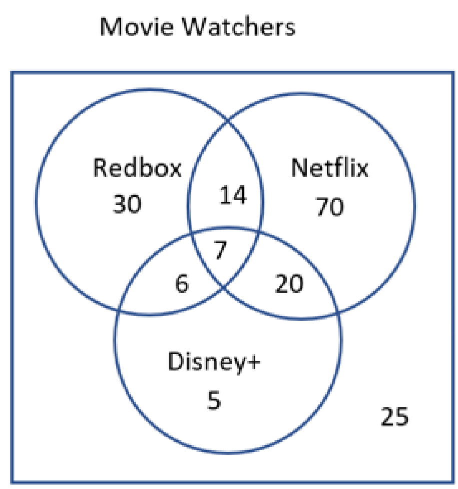
1.2.12.5.

Yes, {1, 3, 5} a subset of the set of odd numbers.

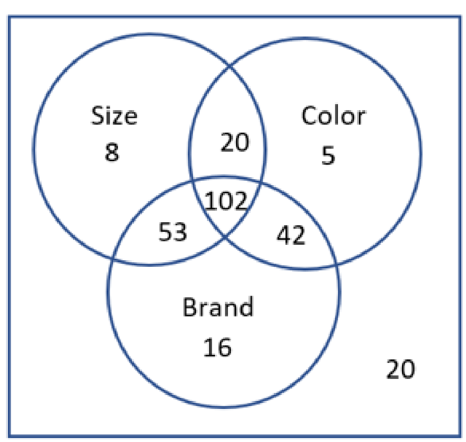
1.2.12.6.

Yes, {A, B, C} a subset of the set of letters of the alphabet.

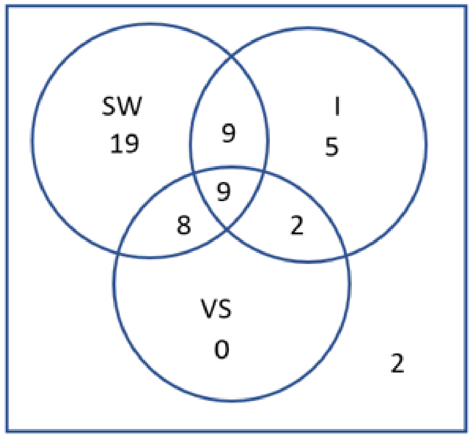
1.2.12.7.



1.2.12.8.



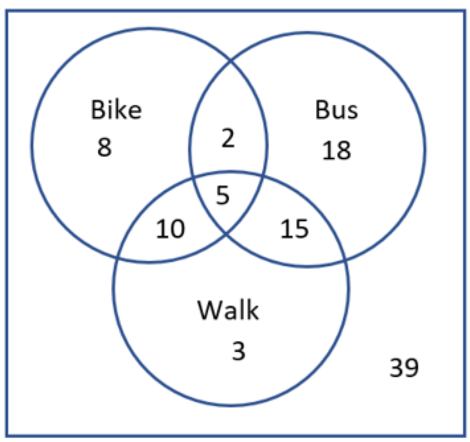
1.2.12.9.



a. Twenty-four students have seen exactly one of these movies.

b. Nineteen students have only seen Star Wars Episode IX.

1.2.12.10.



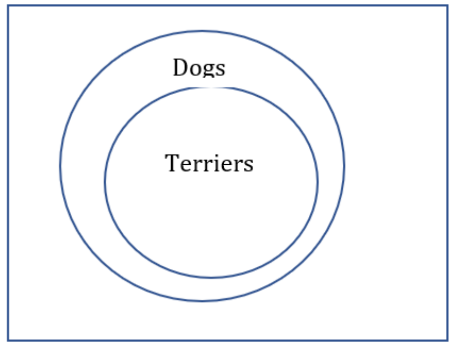
a. Eighteen people only ride the bus.

b. Thirty-nine people don't use any alternate transportation.

1.2.12.11.

a. Terriers and dogs

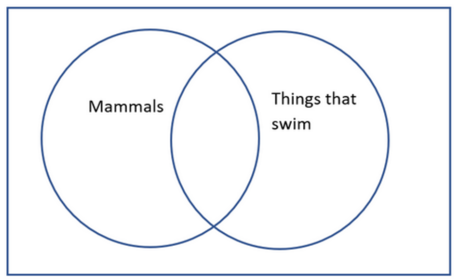
b. Subsets

c. 

1.2.12.12.

a. Mammals and Things that swim

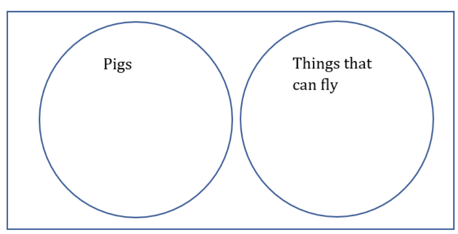
b. Overlapping

c. 

1.2.12.13.

a. Pigs and things that can fly

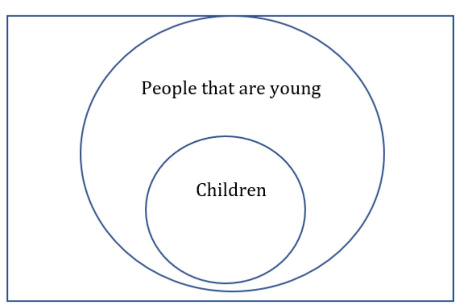
b. Disjoint

c. 

1.2.12.14.

a. Children and people that are young

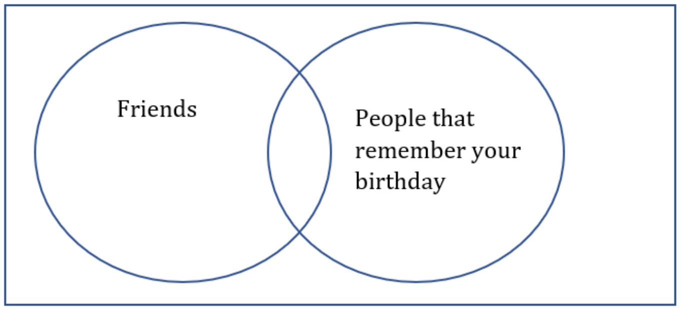
b. Subset

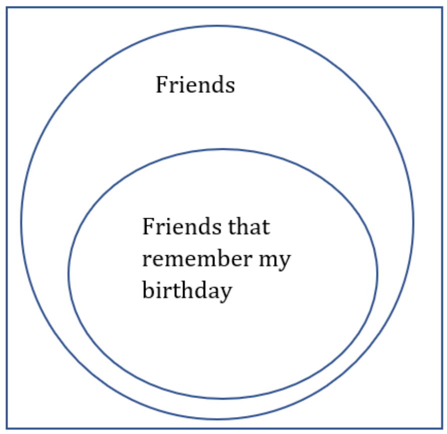
c. 

1.2.12.15.

a. There are at least two ways this can be interpreted. Friends and people that remember your birthday OR friends and friends that remember my birthday.

b. Overlapping or subset (depending on how this is interpreted)

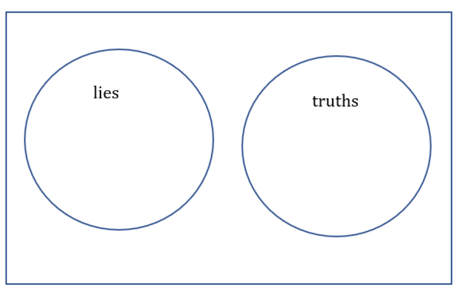
c. 



1.2.12.16.

a. Lies and truths

b. disjoint

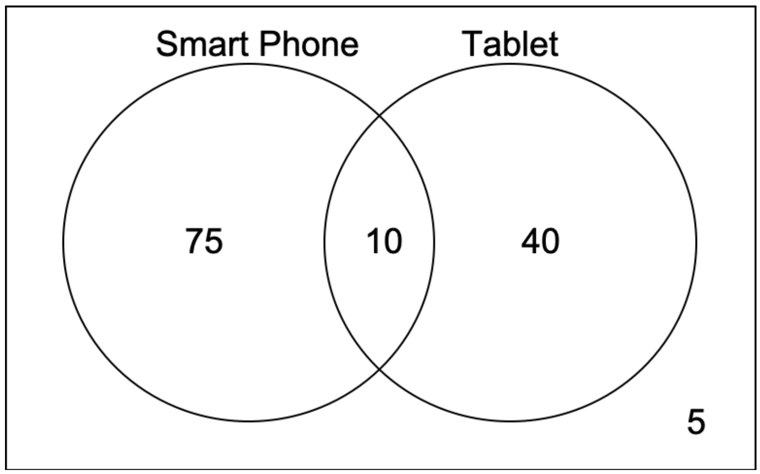
c. 

1.2.12.17.

- Region A: Unemployed non-honor-students who identify as boys
- Region B: Unemployed honor-students who identify as boys
- Region C: Unemployed honor-students who identify as girls or nonbinary
- Region D: Employed non-honor-students who identify as boys

- Region E: Employed honor-students who identify as boys
- Region F: Employed honor-students who identify as girls or nonbinary
- Region G: Employed non-honor-students who identify as girls or nonbinary
- Region H: Unemployed non-honor-students who identify as girls or nonbinary

1.2.12.18.



130 students were surveyed.

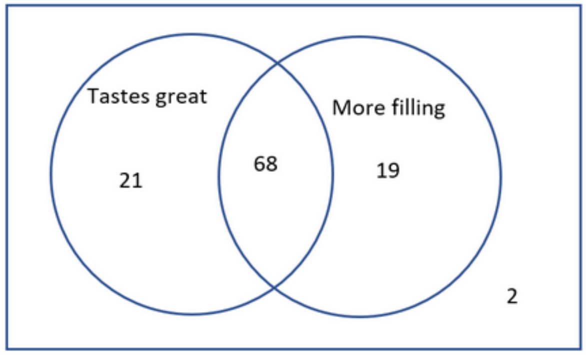
1.2.12.19.

- Students enrolled for keyboard class: 24
- Students enrolled for keyboard class only: 8
- Students didn't enroll at all: 6
- Students took all three classes: 4
- Students enrolled for guitar and drum: 6
- Students enrolled for guitar and drum only: 2

1.2.12.20.

a. Thirty-six coffee drinkers like cream.
b. Fifty-five coffee drinkers like sugar.
c. Thirty-five coffee drinkers like sugar but not cream.
d. Sixteen coffee drinkers like cream but not sugar.
e. Twenty coffee drinkers like cream and sugar.
f. Seventy-one coffee drinkers like cream or sugar.
g. Twenty-nine coffee drinkers like neither cream nor sugar.

1.2.12.21.

a. 

b. Nineteen said “It’s more filling!” but didn’t say “It tastes great!”

c. Forty-two said neither of those things.

d. 108 said “It’s more filling!” or said “It tastes great!”

1.3 · Describing and Critiquing Arguments

1.3.5 · Exercises

1.3.5.1.

a. The argument is inductive.

b. The argument is deductive.

1.3.5.2.

a. The next term is 18.

b. The next term is 39.

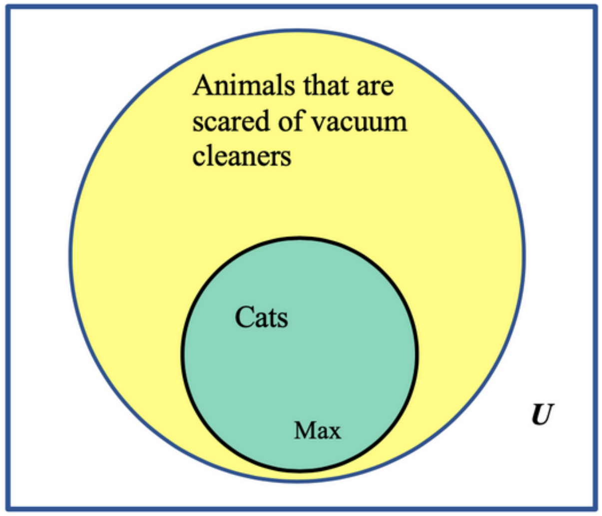
1.3.5.3.

Premise: All cats are scared of vacuum cleaners.

Premise: Max is a cat.

Conclusion: Max must be scared of vacuum cleaners.

This is a deductive argument. It is valid. However, it is not sound because the premise all cats are afraid of vacuum cleaners is false. While many cats are afraid of vacuum cleaners not ALL cats are afraid. There are many videos of cats riding electronic vacuum cleaners.



1.3.5.4.

Premise: Every day for the last year, a plane flew over my house at 2pm.

Conclusion: A plane will always fly over my house at 2pm.

This is an inductive argument. It is a strong argument because a large quantity of data has been collected.

1.3.5.5.

Premise: Kiran's female and nonbinary friends made less than Kiran's male friends.

Conclusion: Women and nonbinary people make less than men.

This is an inductive argument. Kiran did not gather a large diverse sample because they only asked their friends. Therefore, their data has sampling bias. This makes their argument weak.

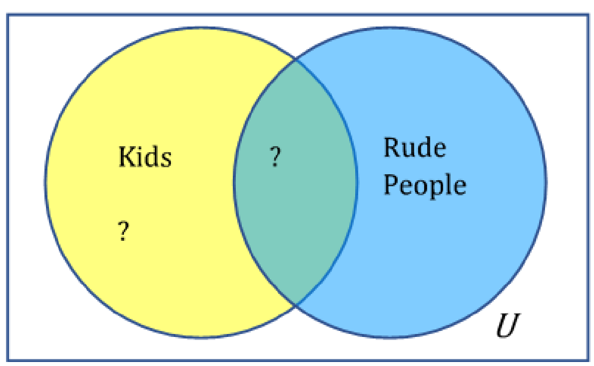
1.3.5.6.

Premise: Some of these kids are rude.

Premise: Jimmy is one of these kids.

Conclusion: Jimmy is rude!

This is a deductive argument. This argument is not valid. Jimmy could be one of these rude kids, he could also be a kid who is not rude. Because the conclusion is not valid, it is also not sound.



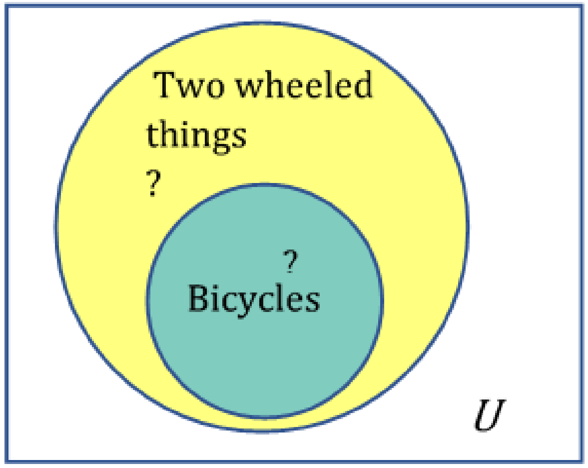
1.3.5.7.

Premise: All bicycles have two wheels.

Premise: My friend's Harley-Davidson motorcycle has two wheels.

Conclusion: It must be a bicycle.

This is a deductive argument. The argument is not valid. Based on the premises, we know that the friend's Harley-Davidson motorcycle has two wheels, but we do not know whethere or not it is a bicycle. Because it is not valid, it is also not sound.



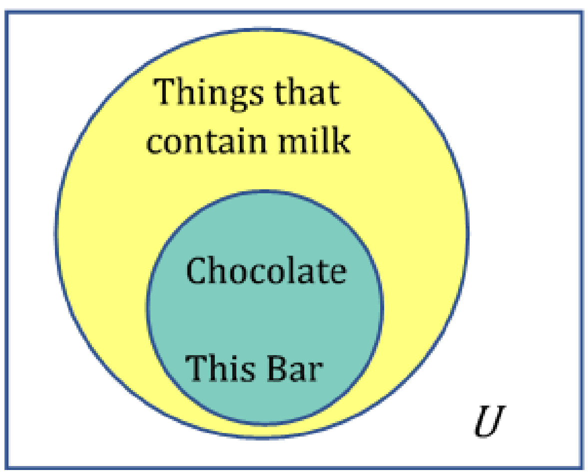
1.3.5.8.

Premise: All chocolate contains milk.

Premise: This bar is made of chocolate.

Conclusion: It must contain milk.

This is a deductive argument. The argument is valid. This bar is in the set of chocolate and chocolate is in the set of things that contain milk. However, this conclusion is not sound because not all chocolate contains milk.



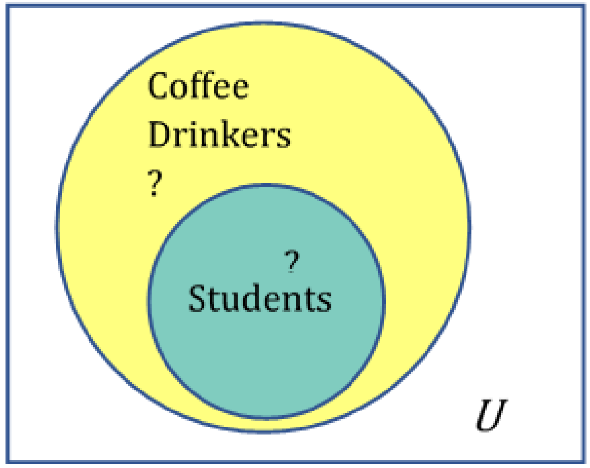
1.3.5.9.

Premise: All students drink a lot of caffeine.

Premise: Brayer drinks a lot of caffeine.

Conclusion: He must be a student.

This is a deductive argument. The argument is not valid. We can not determine if Brayer is a student or not. Because it is not valid it is also not sound.



1.3.5.10.

Premise: Over one year on average, there were 15-35 students present in the cafeteria during the peak hours.

Conclusion: There is going to be between 15 and 35 students in the cafeteria if we go during the peak hours of the day.

This is an inductive argument. This argument is strong because the conclusion is supported by the premises, they gathered a large amount of data over a long period of time on week days.

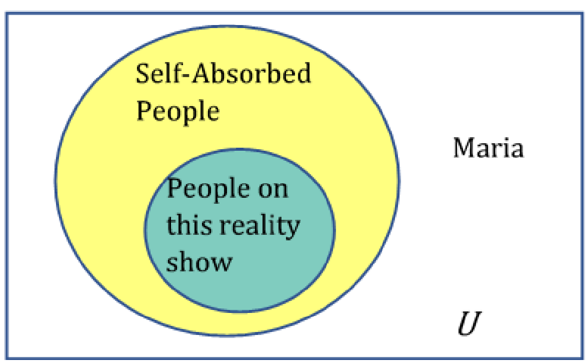
1.3.5.11.

Premise: People on this reality show are self-absorbed.

Premise: Laura is not self-absorbed.

Conclusion: Laura cannot be on this reality show.

This argument is deductive. The argument is valid because Laura is outside of the set of self-absorbed people, so she must also be outside of this set of people on this reality show. Determining if this conclusion is sound is more difficult because determining if someone is self-absorbed is subjective.

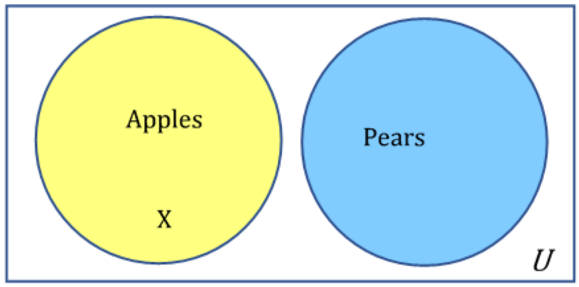


1.3.5.12.

This is a strong inductive argument because we are given specific numbers in which we can find a pattern. Then based on that pattern we can determine that the next term of the sequence would be 11.

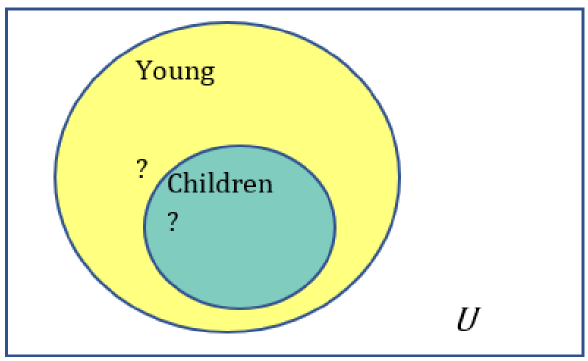
1.3.5.13.

The argument is valid and sound.



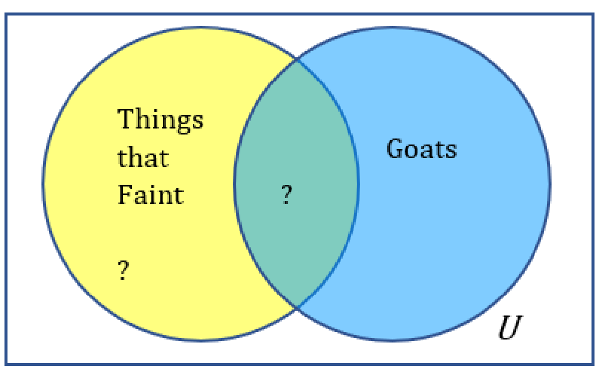
1.3.5.14.

The argument is not valid, and it is not sound.



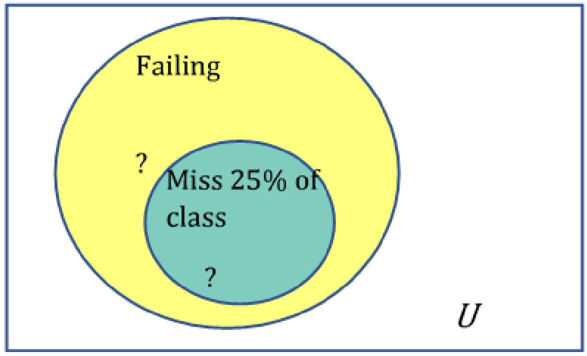
1.3.5.15.

The argument is not valid, and it is not sound.



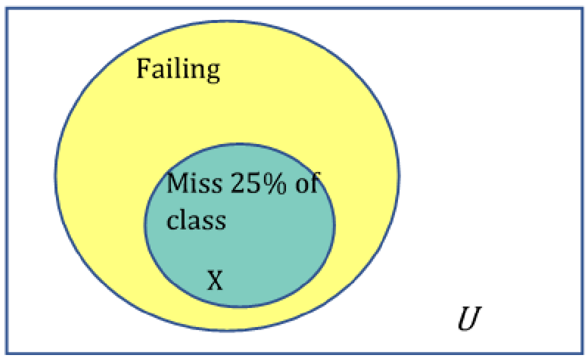
1.3.5.16.

This argument is not valid, and because it is not valid it is also not sound. Claudia could have missed 25% of the classes causing her to fail. She could instead be failing because of low test scores.



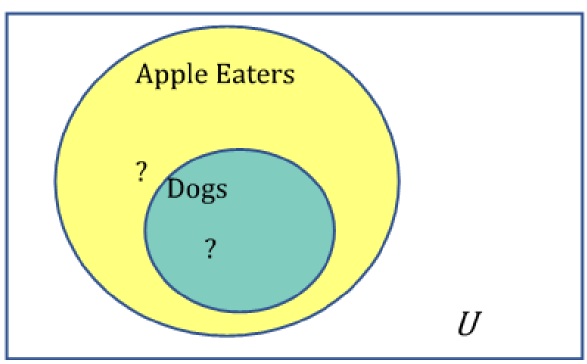
1.3.5.17.

This argument is valid, and it is sound. Because Ethan is in the set of folks who missed 25% of the classes he also falls into the set of failing.



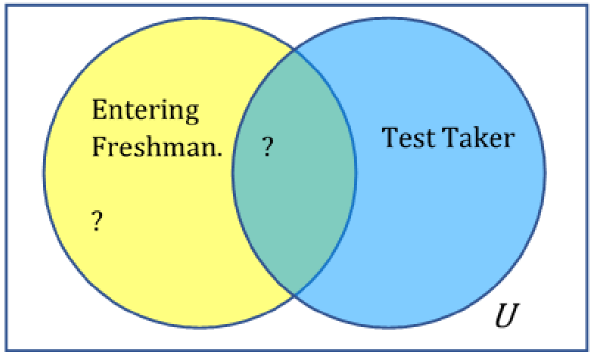
1.3.5.18.

This argument is not valid, and because it is not valid it is also not sound. Mary could be a dog. Mary could also be a worm, horse or human.



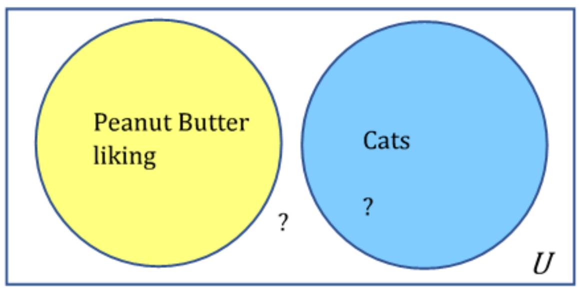
1.3.5.19.

This argument is not valid, and is not sound. We cannot determine if Juan took the test.



1.3.5.20.

This argument is not valid. Because it is not valid, it is also not sound. We cannot determine if Bob is a cat or not.



1.4 · Logical Fallacies

1.4.9 · Exercises

1.4.9.1.

Personal attack

1.4.9.2.

Appeal to authority

1.4.9.3.

Personal Dilemma

1.4.9.4.

False dilemma

1.4.9.5.

Post Hoc

1.4.9.6.

Post Hoc

1.4.9.7.

Personal Attack

1.4.9.8.

Appeal to authority

1.4.9.9.

Appeal to ignorance

1.4.9.10.

Straw person

1.4.9.11.

False dilemma

1.4.9.12.

Appeal to ignorance

1.4.9.13.

Personal attack

1.4.9.14.

Appeal to authority

1.4.9.15.

Appeal to ignorance

1.4.9.16.

Personal attack

1.4.9.17.

Post Hoc

1.4.9.18.

Straw person

1.4.9.19.

Personal Attack

1.4.9.20.

Answers will vary.

1.5 · Chapter 1 Review

· Review Exercises

1.5.1.

a. Yes

b. Yes

c. No

d. No

e. Yes

f. Yes

1.5.2.

a. I don't take public transportation to get to class.

b. I didn't go to a movie on Friday.

c. I want to go golfing today

d. I don't love watching basketball.

e. Breylynn's favorite color is not green.

f. Mirriam is not a theater major.

1.5.3.

a. Inclusive

b. Inclusive

c. Exclusive

d. Exclusive

e. Exclusive

1.5.4.

a. If you do homework, then it helps increase your grade in class.

b. If you ride public transportation, then it will help you save money.

c. If it is a squirrel, then it will bury their food.

d. If you eat to much candy, then you will get sick.

e. If you think you have the flu, then go too the doctor.

1.5.5.

a. I will not buy an iPhone.

b. I do not learn how to use technology fast.

c. I will buy an iPhone or I learn how to use technology fast.

d. I will buy an iPhone and I don't learn how to use technology fast.

e. If I learn to use new technology, then I will buy an iPhone.

1.5.6.

A	B	Not A	Not A and B	Not (Not A and B)
T	T	F	F	T
T	F	F	F	T
F	T	T	T	F
F	F	T	F	T

1.5.7.

A	B	If A, then B
T	T	T
T	F	F
F	T	T
F	F	T

1.5.8.

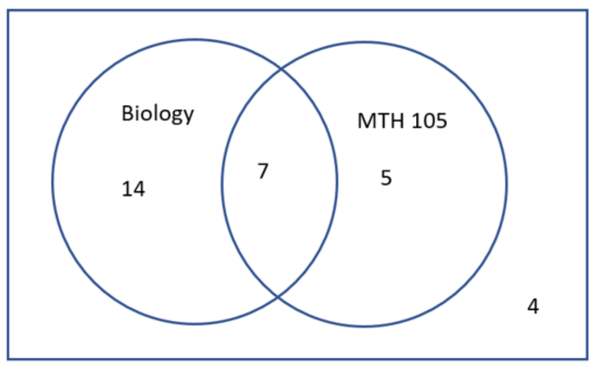
a. 27 people only like cream in their coffee.

b. $13 + 20 = 33$
33 people put sugar in their coffee

c. $27 + 13 + 20 = 60$
60 people put sugar or cream in their coffee.

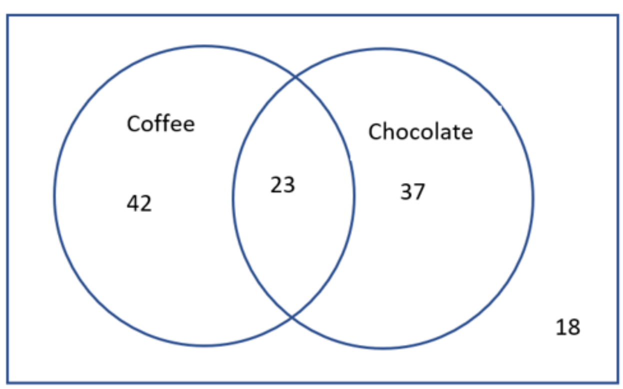
d. $20 + 10 = 30$
30 people don't like cream in their coffee.

1.5.9.

a. 

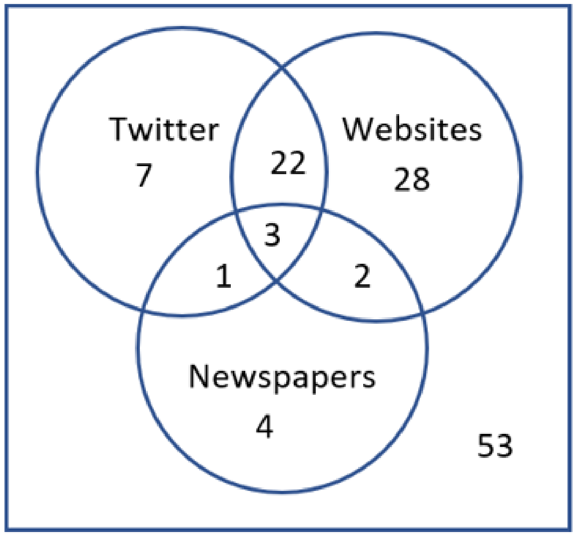
b. $30 - 26 = 4$
4 students are not taking either course.

1.5.10.



1.5.11.

a.



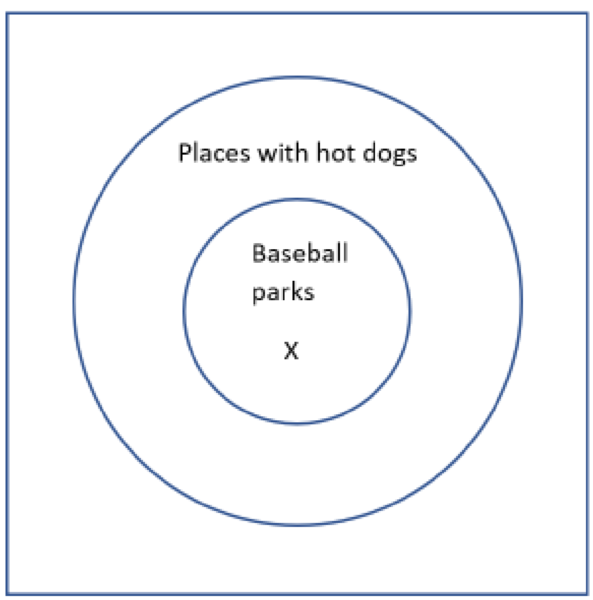
b. $7 + 22 + 28 + 1 + 3 + 2 = 63$.
63 people get thier news from Twitter or a website.

c. 53 people do not get their news from any of the three sources.

d. 7 people use only Twitter.

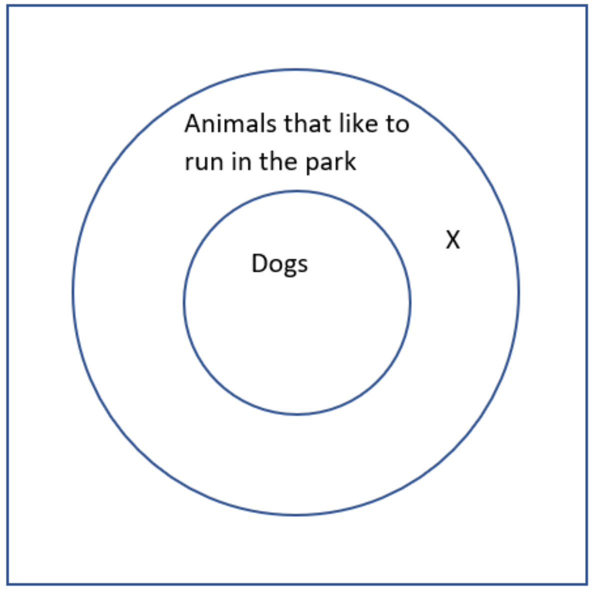
e. $7 + 22 + 28 = 57$.
57 people use Twitter or a website, but not a newspaper.

1.5.12.



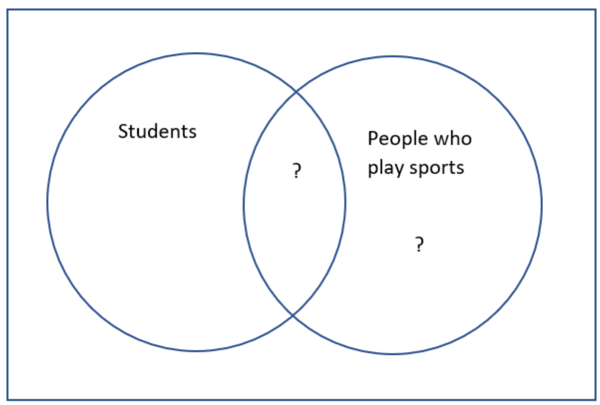
Valid and sound

1.5.13.



Valid, but not sound

1.5.14.



Not valid and not sound

1.5.15.

Premise: Most people find out what's happening on Twitter or Facebook.

Conclusion: It is the most reliable source for news.

Fallacy: Appeal to popularity

1.5.16.

Premise: "Finding the Loch Ness Monster" has yet to provide evidence that Loch Ness exists.

Conclusion: All those sightings are obviously bogus.

Fallacy: Appeal to ignorance

1.5.17.

Premise: Sampson bought a new car, and then he got a traffic ticket for speeding.

Conclusion: Buying the new car must have caused him to speed.

Fallacy: Post Hoc

2 · Financial Math

2.1 · Introduction to Spreadsheets

2.1.6 · Exercises

2.1.6.1.

`=4/7` which gives approximately 0.571429

2.1.6.2.

`=16%` which gives 0.16

2.1.6.3.

`=8+19` which gives 27

2.1.6.4.

`=230-78` which gives 152

2.1.6.5.

`=12*9` which gives 108

2.1.6.6.

`=0.09/52` which gives approximately 0.001731

2.1.6.7.

`=8^3` which gives 512

2.1.6.8.

`=55.75*20%` which gives $11.15

2.1.6.9.

`=7.50/44.50` which gives approximately 0.1685, or approximately 16.85%

2.1.6.10.

See the table below:

Note, the entry in cell B1 is `=15%*A1`. (Columns A and B are given dollar formatting)

	A	B
1	$ 5.00	$ 0.78
2	$ 10.00	$ 1.55
3	$ 15.00	$ 2.33
4	$ 20.00	$ 3.10
5	$ 25.00	$ 3.88
6	$ 30.00	$ 4.65
7	$ 35.00	$ 5.43
8	$ 40.00	$ 6.20
9	$ 45.00	$ 6.98
10	$ 50.00	$ 7.75
11	$ 55.00	$ 8.53
12	$ 60.00	$ 9.30
13	$ 65.00	$ 10.08
14	$ 70.00	$ 10.85
15	$ 75.00	$ 11.63
16	$ 80.00	$ 12.40
17	$ 85.00	$ 13.18
18	$ 90.00	$ 13.95
19	$ 95.00	$ 14.73
20	$ 100.00	$ 15.50
21	$ 105.00	$ 16.28
22	$ 110.00	$ 17.05
23	$ 115.00	$ 17.83
24	$ 120.00	$ 18.60
25	$ 125.00	$ 19.38

2.1.6.11.

a. `=5780.23-5250` which gives $530.23

b. `=530.23/5250` which gives approximately 0.100996, or approximately 10.1%

c. Exactly 200%

d. `=5250*115.5%` which gives exactly $6063.75

2.1.6.12.

a. See the table at the bottom for part a and part b.
After 1 year, the account holds $563.41.

b. See the table at the bottom for part a and part b.
After 2 years, the account holds $634.87.

c. `=A13/A1` which gives $\approx$ 112.6825% growth

d. `=A25/A13` which gives $\approx$ 112.6825% growth (same)

e. Each starting value increases mathematically by a factor of $(1.01)^{12}$ each year, which is approximately 112.6825%. So yes, this pattern must continue indefinitely into future years.

	A
1	$ 500.00
2	$ 505.00
3	$ 510.05
4	$ 515.15
5	$ 520.30
6	$ 525.51
7	$ 530.76
8	$ 536.07
9	$ 541.43
10	$ 546.84
11	$ 552.31
12	$ 557.83
13	$ 563.41
14	$ 569.05
15	$ 574.74
16	$ 580.48
17	$ 586.29
18	$ 592.15
19	$ 598.07
20	$ 604.05
21	$ 610.10
22	$ 616.20
23	$ 622.36
24	$ 628.58
25	$ 634.87

2.1.6.13.

a. `=1000*103%*103%` which gives $1060.90

b. `=1000*(103%)^2` gives the same result of $1060.90, because raising 103% to the second power means the same as multiplying 103% by itself two times.

c. `=1000*(103%)^15` which gives $1557.97 rounded to the nearest cent

d. Refer to the table at the bottom for part d and part e.

Note the entry in cell B3 here is `= B2*103%` and the remaining cells are computed using the fill down feature.

You will have to wait a minimum of 24 full years, in each case, in order for the balance to finally exceed twice the opening deposit amount.

e. Since $(103\%)^{23} < 2 < (103\%)^{24}$, the minimum number of full years until the opening deposit doubles must be the same here, for any positive opening balance that we may choose for this account.

	A	B	C	D	E
1	Year	Balance		Year	Balance
2	0	$ 1,000.00		0	$ 5,000.00
3	1	$ 1,030.00		1	$ 5,150.00
4	2	$ 1,060.90		2	$ 5,304.50
5	3	$ 1,092.73		3	$ 5,463.64
6	4	$ 1,125.51		4	$ 5,627.54
7	5	$ 1,159.27		5	$ 5,796.37
8	6	$ 1,194.05		6	$ 5,970.26
9	7	$ 1,229.87		7	$ 6,149.37
10	8	$ 1,266.77		8	$ 6,333.85
11	9	$ 1,304.77		9	$ 6,523.87
12	10	$ 1,343.92		10	$ 6,719.58
13	11	$ 1,384.23		11	$ 6,921.17
14	12	$ 1,425.76		12	$ 7,128.80
15	13	$ 1,468.53		13	$ 7,342.67
16	14	$ 1,512.59		14	$ 7,562.95
17	15	$ 1,557.97		15	$ 7,789.84
18	16	$ 1,604.71		16	$ 8,023.53
19	17	$ 1,652.85		17	$ 8,264.24
20	18	$ 1,702.43		18	$ 8,512.17
21	19	$ 1,753.51		19	$ 8,767.53
22	20	$ 1,806.11		20	$ 9,030.56
23	21	$ 1,860.29		21	$ 9,301.47
24	22	$ 1,916.10		22	$ 9,580.52
25	23	$ 1,973.59		23	$ 9,867.93
26	24	$ 2,032.79		24	$10,163.97
27	25	$ 2,093.78		25	$10,468.89
28	26	$ 2,156.59		26	$10,782.96
29	27	$ 2,221.29		27	$11,106.45
30	28	$ 2,287.93		28	$11,439.64
31	29	$ 2,356.57		29	$11,782.83
32	30	$ 2,427.26		30	$12,136.31

2.2 · Simple and Compound Interest

2.2.11 · Exercises

2.2.11.1.

$$\begin{aligned} A &= 200 + 200(0.05) \\ &= \$210 \end{aligned}$$

You will have to repay $210.

2.2.11.2.

a.

$$\begin{aligned} A &= 100 + 100(0.03)(0.5) \\ &= \$101.50 \end{aligned}$$

b.

$$\begin{aligned} I &= 101.50 - 100 \\ &= \$1.50 \end{aligned}$$

They paid $101.50 total, $1.50 was interest.

2.2.11.3.

$$\begin{aligned} A &= 200 + 200(0.06)(1.25) \\ &= \$215 \end{aligned}$$

$215 was repaid.

2.2.11.4.

a.

$$\begin{aligned} I &= 1000(0.025)(5) \\ &= \$125 \end{aligned}$$

b.

$$\begin{aligned} A &= 1000 + 1000(0.025)(5) \\ &= \$1125 \end{aligned}$$

The interest earned was $125. This brings the account balance to $1,125.

2.2.11.5.

a.

$$\begin{aligned} A &= 20000 + 20000(0.05)(10) \\ &= \$30,000 \end{aligned}$$

b.

$$\begin{aligned} A &= 20000(1+0.05)^{10} \\ &= \$32,577.89 \end{aligned}$$

The simple interest account would be worth $30,000 and the account that was compounding would be worth $32,577.89 in ten years.

2.2.11.6.

$$\begin{aligned} A &= 4500\left(1+\frac{0.085}{12}\right)^{12*20} \\ &= \$24,485.59 \end{aligned}$$

The account balance is $24,485.59 in 20 years.

Or,

```
=FV(0.085/12,12*20,0,4500)
```

2.2.11.7.

$$\begin{aligned} A &= 1000\left(1+\frac{0.07}{52}\right)^{52*20} \\ &\approx \$4051.38 \end{aligned}$$

The account balance is $4,051.38 in 20 years.

Or,

```
=FV(0.07/52,52*20,0,1000)
```

2.2.11.8.

a.

$$\begin{aligned} A &= 3000\left(1+\frac{0.03}{4}\right)^{4*5} \\ &\approx \$3,483.55 \end{aligned}$$

The future value is $3,483.55.

Or,

```
=FV(0.03/4,4*5,0,3000)
```

b.

$$\begin{aligned} I &= 3483.53 - 3000 \\ &= \$483.54 \end{aligned}$$

The interest earned is $483.54.

c.

$$\frac{483.54}{3483.54} \approx 0.1388 \text{ or } 13.88\%$$

13.88% of the balance is interest.

2.2.11.9.

a.

$$\begin{aligned} A &= 300\left(1+\frac{0.05}{1}\right)^{1*10} \\ &\approx \$488.67 \end{aligned}$$

There will be \$488.67 in the account in 10 years.

Or,

```
=FV(0.05/1,1*10,0,300)
```

b.

$$\begin{aligned} I &= 488.67 - 300 \\ &= \$188.67 \end{aligned}$$

\$188.67 of the balance will be interest.

c.

$$\frac{188.67}{488.67} \approx 0.3861 \text{ or } 38.61\%$$

The interest makes up 38.61% of the balance.

2.2.11.10.

a.

$$\begin{aligned} A &= 2000\left(1+\frac{0.03}{12}\right)^{12*20} \\ &\approx \$3,641.51 \end{aligned}$$

The account will have \$3,641.51 in 20 years

Or,

```
=FV(0.03/12,12*20,0,2000)
```

b.

$$\begin{aligned} I &= 3641.51 - 2000 \\ &= \$1641.51 \end{aligned}$$

The interest will be \$1641.51.

c.

$$\frac{1641.51}{3641.51} \approx 0.4508 \to 45.08\%$$

The interest is 45.08% of the balance.

d.

$$\frac{2000}{3641.51} \approx 0.5492 \to 54.92$$

The principal is 54.92% of the balance.

2.2.11.11.

a.

$$\begin{aligned} A &= 10000\left(1+\frac{0.04}{52}\right)^{52*25} \\ &\approx \$27,172.37 \end{aligned}$$

The balance is \$27,172.37.

Or,

```
=FV(0.04/52,52*25,0,10000)
```

b.

$$\begin{aligned} I &= 27172.37-10000 \\ &= \$17,172.37 \end{aligned}$$

The interest is \$17,172.37.

c.

$$\frac{17172.37}{27172.37} \approx 0.632 \rightarrow 63.2\%$$

The percent that is interest is 63.2%.

d.

$$\frac{10000}{27172.37} \approx 0.368 \rightarrow 36.8\%$$

The percentage that is the principal is 36.2%.

2.2.11.12.

$$\begin{aligned} P &= \frac{6000}{\left(1+\frac{0.06}{12}\right)^{12*8}} \\ &\approx \$3,717.14 \end{aligned}$$

The principal required would be \$3717.14

Or,

```
=PV(0.06/12,12*8,0,6000)
```

2.2.11.13.

$$\begin{aligned} P &= \frac{20000}{\left(1+\frac{0.05}{4}\right)^{4*4}} \\ &\approx \$16,394.79 \end{aligned}$$

The principal required would be \$16,394.79

Or,

`=PV(0.05/4,4*4,0,20000)`

2.2.11.14.

a.

$$A_{Breylan} = 1200\left(1 + \frac{0.046}{52}\right)^{52*15}$$
$$\approx \$2,391.73$$

Or, `=FV(0.046/52,15*52,0,1200)`

$$A_{Angad} = 1200\left(1 + \frac{0.0455}{52}\right)^{52*15}$$
$$\approx \$2,373.87$$

Or, `=FV(0.0455/52,15*52,0,1200)`

Breylan has an account balance of $2,391.73 and Angad has a balance of $2,373.87

b.

$$A_{Breylan} = 1200\left(1 + \frac{0.046}{52}\right)^{52*30}$$
$$\approx \$4,766.97$$

Or, `=FV(0.046,52,30*52,0,1200)`

$$A_{Angad} = 1200\left(1 + \frac{0.0455}{52}\right)^{52*30}$$
$$\approx \$4,696.06$$

Or, `=FV(0.0455/52,30*52,0,1200)`

Breylan has an account balance of $4,766.97 and Angad has a balance of $4,696.06

c. Breylan `=EFFECT(0.046,52)`, which gives 4.71%

Angad `=EFFECT(0.0455,52)`, which gives 4.65%

Breylan has an effective rate of 4.71% and Angad has an effective rate of 4.65%.

2.2.11.15.

a. Bill `=EFFECT(0.0375,12)` $= 3.82$

Or, `=FV(0.0375/12,5*12,0,6700)`

$$A = 6500\left(1 + \frac{0.038}{1}\right)^{1*5}$$
$$\approx \$7,832.49$$

Or, `=FV(0.038,5,0,6500)`

The account balances are $8,079.38 and $7,832.49. So, Bill's balance is higher.

2.2.11.16.

a. `=EFFECT(0.0345,4)`= 3.49% and `=EFFECT(0.034,365)`= 3.46%. The effective rates are 3.49% and 3.46%.

b.

$$\begin{aligned} A &= 5000\left(1+\frac{0.0345}{4}\right)^{4*10} \\ &\approx \$7,049.51 \end{aligned}$$

Or, `=FV(0.0345/4,10*4,0,5000)`

$$\begin{aligned} A &= 5000\left(1+\frac{0.034}{365}\right)^{365*10} \\ &\approx \$7,024.63 \end{aligned}$$

Or, `=FV(0.034/365,10*365,0,5000)`

The account balances are \$7,049.51 and \$7,024.63.

2.2.11.17.

a.

$$\begin{aligned} A &= 2500e^{0.04*10} \\ &\approx \$3,729.56 \end{aligned}$$

Or, `2500*EXP(0.04*10)`

The account balance is \$3,729.56.

b.

$$\begin{aligned} I &= 3729.56 - 2500 \\ &= \$1229.56 \end{aligned}$$

The interest earned is \$1,229.56.

c.

$$\frac{1229.56}{3729.56} \approx 0.3297 \to 32.97\%$$

32.97% of the balance is interest.

2.2.11.18.

a.

$$\begin{aligned} A &= 1000e^{0.0575*15} \\ &\approx \$2,369.08 \end{aligned}$$

Or, `=1000*EXP(0.0575*15)`

The account balance is \$2,369.08

b.

$$\begin{aligned} I &= 2368.08 - 1000 \\ &= \$1,368.08 \end{aligned}$$

The interest earned is $1,368.08

c.

$$\frac{1368.08}{2368.08} \approx 0.5777 \to 57.77\%$$

The interest is 57.77% of the account balance.

2.2.11.19.

a.

$$\begin{aligned} A &= 5000e^{0.045*5} \\ &\approx \$6,261.61 \end{aligned}$$

Or, `=5000*EXP(0.045*5)`

The account balance is $6,261.61.

b.

$$\begin{aligned} I &= 6261.61 - 5000 \\ &= \$1,261.61 \end{aligned}$$

The interest earned is $1,261.61

c.

$$\frac{1261.61}{6261.61} \approx 0.2015 \to 20.15\%$$

The interest is 20.15% of the balance.

2.2.11.20.

a.

$$\begin{aligned} A_{You} &= 10000e^{0.055*10} \\ &\approx \$17,332.53 \end{aligned}$$

Or, `=10000*EXP(0.055*10)`

$$\begin{aligned} A_{Friend} &= 10000(1 + \frac{0.055}{1})^{1*10} \\ &\approx \$17,081.44 \end{aligned}$$

Or, `=FV(0.055,10,0,10000)`

Your balance is $17,332.53 and your friend's balance is $17,081.44.

b. The difference is:

$$17332.53 - 17081.44 = \$251.09$$

You have $251.09 more than your friend.

2.3 · Savings Plans

2.3.5 · Exercises

2.3.5.1.

a. $\frac{250[(1+\frac{0.065}{12})^{12*35}-1]}{(\frac{0.065}{12})}$

   Or, `FV(0.065/12,12*35,250)`

   In 35 years you will have $400,079.05 in your retirement plan.

b. $400079.05\breve{}12 * 35 * 250$. You will have earned $295,079.05 in interest.

c. $295079.05/400079.05$. The final balance will be about 73.8% interest.

2.3.5.2.

a. $\frac{75[(1+\frac{0.085}{52})^{52*40}-1]}{(\frac{0.085}{52})}$

   Or, `=FV(0.085/52,52*40,75)`

   In 40 years you will have $1,325,130.09 in your retirement plan.

b. $1325130.09\breve{}52 * 40 * 75$. You will have earned $1,169,130.09 in interest.

c. $1169130.09/1325130.09$. The final balance will be about 88.2% interest.

2.3.5.3.

a. $\frac{750[(1+\frac{0.0775}{4})^{4*30}-1]}{(\frac{0.0775}{4})}$

   Or, `=FV(0.0775/4,4*30,750)`

   In 30 years you will have $348,456.10 in your retirement plan.

b. $348456.1\breve{}4 * 30 * 750$. You will have earned $258,456.10 in interest.

c. $258456.1/348456.1$. The final balance will be about 74.2% interest.

2.3.5.4.

a. $\frac{20[(1+\frac{0.0235}{365})^{365*25}-1]}{(\frac{0.0235}{365})}$

   Or, `=FV(0.0235/365,365*25,20)`

   In 25 years you will have $248,339.80 in your retirement plan.

b. $248339.8\breve{}365 * 25 * 20$. You will have earned $65,839.80 in interest.

c. $65839.8/248339.8$. The final balance will be about 26.5% interest.

2.3.5.5.

a. In 5 years: $\frac{130[(1+\frac{0.09}{12})^{12*5}-1]}{(\frac{0.09}{12})}$

   Or, `=FV(0.09/12,12*5,130)`

   In 25 more years: $9805.14(1 + \frac{0.09}{12})^{12*25}$

   `=FV(0.09/12,12*25,0,9805.14)`

   Your final balance will be $92,250.82.

b. $92250.82˘130 * 5 * 12$. You will earn \$84,450.82 in interest.

c. $84450.82/92250.82$. The final balance will be about 91.5% interest.

2.3.5.6.

a. $\frac{200[(1+\frac{0.05}{12})^{12*20}-1]}{(\frac{0.05}{12})}$

Or, `=FV(0.05/12, 12*10, 200)`

After 10 years the balance will be \$31,056.46.

b. In 20 more years: $31056.46(1+\frac{0.05}{12})^{12*20}$

Or, `=FV(0.05/12,12*20,0,31056.46)`

The final balance will be \$84,245.00.

c. $84245˘200 * 12 * 10$. You will earn \$60,245 in interest.

d. $60245/84245$. Your final balance will be about 71.5%. interest.

2.3.5.7.

$\frac{3500(\frac{0.038}{12})}{(1+\frac{0.038}{12})^{30}-1}$

`=PMT(0.038/12,30,0,3500)`

You should deposit \$111.40 each month.

2.3.5.8.

$\frac{3000(\frac{0.065}{4})}{(1+\frac{0.065}{4})^{4*2}-1}$

Or, `=PMT(0.065/4,4*2,0,3000)`

You should deposit \$354.19 each quarter.

2.3.5.9.

$\frac{450000(\frac{0.06}{12})}{(1+\frac{0.06}{12})^{12*30}-1}$

Or, `=PMT(0.06/12,12*30,0,450000)`

Jamie needs to deposit \$447.98 each month.

2.3.5.10.

$\frac{500000(\frac{0.1}{4})}{(1+\frac{0.1}{4})^{4*40}-1}$

Or, `=PMT(0.1/4,4*40,0,500000)`

Lashonda should deposit \$245.20 each quarter.

2.3.5.11.

a. Jose: $55000(1+\frac{0.056}{12})^{12*25}$

Or, `=FV(0.056/12,12*25,0,55000)`

Jose's partner: $\frac{375[(1+\frac{0.056}{12})^{12*25}-1]}{(\frac{0.056}{12})}$

Or, `=FV(0.056/12,12*25,375)`

Jose will have \$222,310.85 and his partner will have \$244,447.68.

b. Jose: $222310.85˘55000$

Jose's partner: $244447.68˘375 * 12 * 25$

Jose will earn $167,310.85 and his partner will earn $131,947.68 in interest.

c. Jose: $167310.85/222310.85$

Jose's partner: $131947.68/244447.68$

Jose's final balance will be about 75.3% interest and Jose's partner's final balance will be about 54.0% interest.

2.3.5.12.

a. Akiko: $45000(1+\frac{0.078}{12})^{12*30}$

Or, `=FV(0.078/12,12*30,0,45000)`

Her spouse: $\frac{200[(1+\frac{0.078}{12})^{12*30}-1]}{(\frac{0.078}{12})}$

Or, `=FV(0.078/12,12*30,200)`

Akiko will have $463,631.61 and her spouse will have $286,243.83.

b. Akiko: $463631.61\check{}45000$

Her spouse: $286243.83\check{}200*12*30$

Akiko will earn $418,631.61 and her spouse will earn $214,243.83 in interest.

c. Akiko: $418631.61/463631.61$

Her spouse: $214243.83/286243.83$

Akiko's final balance will be about 90.3% interest and her spouse's final balance will be about 74.8% interest.

2.3.5.13.

a. $1000(1+\frac{0.045}{12})^{12*10}+\frac{100[(1+\frac{0.045}{12})^{12*10}-1]}{\frac{0.045}{12}}$

Or, `=FV(0.045/12,12*10,100,1000)`

Sylvin will have a final balance of $16,686.80.

b. $16686.8\check{}(1000+100*12*10)$. Sylvin will earn $3,686.80 in interest.

c. $3686.8/16686.8$. The final balance will be about 22.1% interest.

2.3.5.14.

a. $5000(1+0.0235)^{20}+\frac{1000[(1+0.0235)^{20}-1]}{0.0235}$

`=FV(0.0235,20,1000,5000)`

Elena's final balance will be $33,119.05.

b. $33119.05\check{}(5000+1000*20)$. Elena will earn $8,119.05 in interest.

c. $8119.05/33119.05$. The final balance will be about 24.5% interest.

2.3.5.15.

a. $\frac{100[(1+\frac{0.04}{12})^{12*25}-1]}{\frac{0.04}{12}}$

Or, `=FV(0.04/12,12*25,100)`

Vanessa will have $51,412.95 when she turns 65.

b. $\frac{100[(1+\frac{0.04}{12})^{12*40}-1]}{\frac{0.04}{12}}$

`=FV(0.04/12,12*40,100`

Vanessa would have $118,196.13 if she had started saving when she was 25.

2.3.5.16.

a. $\frac{1000[(1+0.05)^{20}-1]}{0.05}$

Or, =FV(0.05,20,1000)

Chris will have $33,065.95 after 20 years.

b. $\frac{1000[(1+0.05)^{40}-1]}{0.05}$

Or, =FV(0.05,40,1000)

Chris will have $120,799.77 after 40 years.

2.3.5.17.

a. $\frac{50[(1+\frac{0.035}{52})^{52*18}-1]}{\frac{0.035}{52}}$

Or, =FV(0.035/52,52*18,50)

They will have saved $65,164.37 after 18 years.

b. $\frac{65164.37}{(1+\frac{0.035}{52})^{52*18}}$

Or, =PV(0.035/52,52*18,0,65164.37)

They would have needed an deposit of $34,713.37.

2.3.5.18.

a. $\frac{100[(1+\frac{0.045}{12})^{12*20}-1]}{\frac{0.045}{12}}$

Or, =FV(0.045/12,12*10,100)

Elisa will have $15,119.81 in 10 years.

b. $\frac{15119.81}{(1+\frac{0.045}{12})^{12*10}}$

Or, =PV(0.045/12,12*10,0,15119.81)

She would have needed an initial deposit of $9,648.93.

2.4 · Loan Payments

2.4.9 · Exercises

2.4.9.1.

a. $P = \frac{700\left(1-\left(1+\frac{5.5\%}{12}\right)^{-12\times 30}\right)}{\frac{5.5\%}{12}}$, which gives $P \approx \$123,285.23$

or =PV(5.5%/12,12*30,-700) [Note 700 is entered as negative, to signify a payment]

b. $700 * 12 * 30$ dollars, or $252,000.00 in total payments to the loan company

c. Interest will be the difference between the total payments, and the amount borrowed. So the interest on this loan is $\$252,000.00 - \$123,285.23 = \$128,714.77$.

2.4.9.2.

a. $P = \frac{250\left(1-\left(1+\frac{7\%}{12}\right)^{-12\times 5}\right)}{\frac{7\%}{12}}$, which gives $P \approx \$12,625.50$

or =PV(7%/12,12*5,-250)

b. $250 * 12 * 5$ dollars, or $15,000.00 in total payments to the loan company.

c. Interest will be the difference between the total payments, and the amount borrowed. So the interest on this loan is $\$15,000.00 - \$12,625.50 = \$2,374.50$.

2.4.9.3.

$d = \frac{25000\left(\frac{2\%}{12}\right)}{\left(1-\left(1+\frac{2\%}{12}\right)^{-12\times 4}\right)}$, which gives $d \approx \$542.38$

or `=PMT(2%/12,48,25000)`

2.4.9.4.

$d = \frac{12000\left(\frac{3\%}{12}\right)}{\left(1-\left(1+\frac{3\%}{12}\right)^{-12\times 4}\right)}$, which gives $d \approx \$265.61$

or `=PMT(3%/12,48,12000)`

The interest will be the total amount paid, minus the amount of the loan. So the interest here is ($\$265.61 * 48)\check{}\$12,000 = \$749.28$

2.4.9.5.

a. The loan amount will be 90% of \$200,000.00

$$\begin{aligned} &= (0.9 * \$200,000.00) \\ &= \$180,000.00 \end{aligned}$$

b. $d = \frac{180000\left(\frac{5\%}{12}\right)}{\left(1-\left(1+\frac{5\%}{12}\right)^{-12\times 30}\right)}$, which gives $d \approx \$966.28$

   or `=PMT(5%/12,12*30,180000)`

c. $d = \frac{180000\left(\frac{6\%}{12}\right)}{\left(1-\left(1+\frac{6\%}{12}\right)^{-12\times 30}\right)}$, which gives $d \approx \$1,079.19$

   or `=PMT(6%/12,12*30,180000)`

2.4.9.6.

a. $d = \frac{270000\left(\frac{6.5\%}{12}\right)}{\left(1-\left(1+\frac{6.5\%}{12}\right)^{-12\times 30}\right)}$, which gives $d \approx 1,706.58$

   or `=PMT(6.5%/12,12*30,270000)`

b. The total of all loan payments will be $(12 * 30 * \$1,706.58) = \$614,368.80$.

   So the total interest paid will be $\$614,368.80\check{}\$270,000.00 = \$344,368.80$.

c. $(\$344,368.80/\$614,368.80)$, or $\approx 56.0525\%$

2.4.9.7.

First, we need to find out the amount of the monthly payments for this loan.

$d = \frac{24000\left(\frac{3\%}{12}\right)}{\left(1-\left(1+\frac{3\%}{12}\right)^{-12\times 5}\right)}$, which gives $d \approx \$431.25$

or `=PMT(3%/12,12*5,24000)`

The amount still owed three years later, is the present value of the two years of remaining payments on the loan.

$P = \frac{\left(1-\left(1+\frac{3\%}{12}\right)^{-12\times 2}\right)}{\frac{3\%}{12}}$, which gives $P \approx \$10,033.45$

or `=PV(3%/12,12*2,-431.25)`

2.4.9.8.

First, we need to find out the amount of the monthly payments for this loan.

$d = \frac{120000\left(\frac{6\%}{12}\right)}{\left(1-\left(1+\frac{6\%}{12}\right)^{-12\times 30}\right)}$, which gives $d \approx \$719.46$

or `=PMT(6%/12,12*30,120000)`

The amount still owed fifteen years later, is the present value of the fifteen years of remaining payments on the loan.

$P = \frac{719.46\left(1-\left(1+\frac{6\%}{12}\right)^{-12\times 15}\right)}{\frac{6\%}{12}}$, which gives $P \approx \$85,258.54$

or `=PV(6%/12,12*15,-719.46)`

2.4.9.9.

`=PV(4%/12,12*30,950)` which gives $198,988.18.

Now add $20,000 to this balance, which gives $218,988.18.

2.4.9.10.

a. `=PMT(2.95%/12,12*15,180000,0)` which gives $1,238.72.

   You cannot afford this home.

b. Using trial-and-error: The minimum number of additional years is 8. There is a 23 year loan period, 3.35% APR, and $936.24 monthly payment.

2.4.9.11.

a. `=PMT(5%,10,100000,0)` which gives $12,950.46.

b. The total payments will be $129,504.60.

   The interest is the amount over the $100,000 initial investment, or $29,504.60.

   The percentage of the total payment sum representing interest will be $100*(29,504.60/129,504.60)\%$, or approximately 22.7827%.

2.4.9.12.

a. `=PV(3.5%/4,4*12,1500)` which gives $58,586.02.

b. `=PMT(3.5%/4,4*12,0,100000)` which gives $1,685.34.

c. `=PMT(3.5%/4,4*12,-8000,100000)` which gives $1,480.51.

   (Note 8000 is input as negative, to signify a payment.)

   Note this answer can also be found by subtracting the future value of $8000 here, which equals $12,153.47, from the required ending balance of $100,000 (which leaves $87,846.47) and then using the PMT function:

   `=PMT(3.5%/4,4*12,0,100000-12153.47)` which again gives $1,480.51.

   This can also be done in nested fashion (be careful to add the negative FV quantity): `=PMT(3.5%/4,4*12,0,100000+FV(3.5`

2.4.9.13.

a. `=PMT(4.5%/12,12*30,250000)` which gives $1,266.71.

b. This will be the present value of the remaining 240 loan payments:

   `=PV(4.5%/12,12*20,1266.71)` which gives $200,223.07.

c. This will be the present value of the remaining 120 loan payments:

`=PV(4.5%/12,12*10,1266.71)` which gives \$122,223.99.

d. At the beginning of the repayment period, most of each payment goes to interest (thus the loan balance reduces very slowly at first). Over time, more of each payment shifts to principal, and less to interest. At the end of the loan repayment period, nearly all the payment is going to principal.

2.4.9.14.

`=20000*EXP(0.069*8)` = \$34,734.46. Keisha will have \$34,734.46 in 8 years.

2.4.9.15.

`=FV(0.035/12,3*12,200,0)` = \$7,579.95. Paul will have \$7,579.95 in 3 years.

2.4.9.16.

`=PV(0.05/1,6*1,0,30000)` = \$22,386.46. Sol should deposit \$22,386.46 to have \$30,000 in 6 years.

2.4.9.17.

`=PV(0.028/12,4*12,100,0)` = \$4,535.96. Miao can finance \$4,535.96 in equipment to have a monthly loan payment of \$100 for 4 years.

2.4.9.18.

`=PMT(0.041/12,2*12,0,5000)` = \$200.26. You would need to save \$200.26 every month to have \$5,000 in 2 years.

2.4.9.19.

`=PMT(0.043/12,30*12,364500,0)` = \$1,803.81. Their mortgage payment would be \$1,803.81.

2.4.9.20.

`=500+500*.03*2` = \$530. Your sister should pay you back \$530 in 2 years.

2.4.9.21.

For the first plan: `=FV(0.048/12,20*12,150,0)` = \$60,251.26. Zahid would have \$60,251.26 if he puts in \$150 per month for 20 years.

For the second plan: `=FV(0.048/12,10*12,300,0)` = \$46,089.59. Zahid would only have \$46,089.59 if he waited and put in \$300 per month for 10 years.

2.5 · Income Taxes

2.5.14 · Exercises

2.5.14.1.

A credit decreases your bill more. It decreases your bill by the full amount of the credit. A deduction only decreases your tax bill by a percentage.

2.5.14.2.

The amount of taxes owed is decreased by \$500.

2.5.14.3.

The amount of taxes owed is decreased by \$60 because 12% of \$500 is \$60.

2.5.14.4.

No, you must choose to take either the standard deduction or itemized deductions.

2.5.14.5.

Yes, you can make adjustments and take a deduction. Adjustments to your income happen before deductions.

2.5.14.6.

If you decide to take the standard deduction you determine that itemizing would not save you more money.

2.5.14.7.

She should itemize because itemizing reduces her taxable income by $13,000. The standard deduction would have reduced her taxable income by $12,000.

2.5.14.8.

You can either file as married filing jointly or married filing separately.

2.5.14.9.

No, he should not be concerned. Only the $1000 will be taxed at 22%. The rest of his income will be taxed at a lower level.

2.5.14.10.

The first $9,525 will be taxed at 10% resulting in $952.50 owed in taxes.

2.5.14.11.

There is a total of $29,175 between $9,525 and $38,700. 12% of $29,175 is $3,501.

2.5.14.12.

$80,000 falls into the third tax bracket. The first two brackets resulted in $952.50 and $3,501. The third bracket has $41,300 being taxed at the higher rate of 22%. This results in a tax of $13,539.50.

$4453.50 + 0.22 * (80000 - 38700) = \$13,539.50$

2.5.14.13.

$16,589 in taxes minus $13,456 for withholdings, and they can claim and $2,500 in credits. This leaves them owing $633.

2.5.14.14.

$7,589 in taxes minus $6,456 for withholdings and education credit of $1,980. This leaves her refunded $847.

2.5.14.15.

a. Take the income minus the adjustments $125000 - 5600 = \$119,400$. Their adjusted gross income is $119,400.

b. They should take the standard deduction because itemizing saves them less.

c. Taxable income: $119400 - 24000 = 95400$

   Taxes owed: $8907 + 0.22 * (95400 - 77400) = \$12,867$

   They owe $12,867 in taxes.

d. $12867 - 15000 = -\$2,133$ Take the taxes owed minus credits and withholdings. They will receive a refund for $2,133.

2.5.14.16.

SINGLE CASE:

- Francis: Taxable Income $35000 - 7000 - 12000 = \$16,000$
- Tax from Table: $952.5 + 0.12 * (16000 - 9525) = \$1,729.50$

- Owed/Refund: $1729.50\breve{}14000 - 4000 = -\$16,270.50$

- Edward: Taxable Income $40000 - 3000 - 12000 = \$25,000$
- Tax from Table: $952.50 + 0.12 * (25000 - 9525) = \$2,809.50$
- Owed/Refund: $2809.50 - 5500 - 5000 = -\$7,690.50$

Altogether, they will get a $16270.50 + 7690.50 = \$23,961$ refund.

MARRIED CASE:

- Taxable income: $(35000 + 40000)\breve{}(7000 + 3000)\breve{}24000 = \$41,000$
- Tax from Table: $1905 + 0.12 * (41000 - 19050) = \$4,539$
- Owed/Refund: $4539\breve{}(14000 + 5500)\breve{}(4000 + 5000) = -\$23,961$

Same either way! No marriage penalty unless it takes away credits.

2.5.14.17.

a. Gross Income: $76000 + 750 = \$76,750$
b. Adjusted Gross Income: $76750 - 25000 = \$51,750$
c. She should take itemized deductions since they are greater than the standard deduction for head of household.
d. Taxable Income: $51750 - 19600 = \$32,150$
e. Tax from Table: $1360 + 0.12 * (32150 - 13600) = \$3,586$. Owed/Refund: $3586\breve{}4200\breve{}6300 = -\$6,914$
   Janice will receive a refund for $6,914.

2.6 · Chapter 2 Review

· Review Exercises

2.6.1.

a. $I = 1525(0.056)(14)$
   $A = 1525 + 1525(0.056)(14)$
   The interest is $1,195.6 and balance is $2,720.6.
b. $A = 1525(1 + \frac{0.056}{4})^{4*14}$
   Or, `=FV(0.056/4, 4*14, 0, 1525)`
   The interest is $1,796.97 and the balance is $3,321.97.
c. $A = 1525(1 + \frac{0.056}{52})^{52*14}$
   Or, `=FV(0.056/52, 52*14, 0, 1525)`
   The interest is $1,813.67 and the balance is $3,338.67.
d. $A = 1525e^{0.056*14}$
   Or, `=1525*EXP(0.056*14)`
   The interest is $1,815.08 and the balance is $3,340.08.

2.6.2.

$I = 2300(0.15)(3)$

$A = 2300 + 2300(0.15)(3)$

The interest is $1,035 and the amount paid is $3,335.

2.6.3.

$A = 25000(1 + \frac{0.065}{4})^{4*15}$

$I = 65761.77 - 25000 = 40761.77$

Or, `=FV(0.065/4, 4*15, 0, 25000)`

The interest is $40,761.77 and the amount paid is $65,761.77.

2.6.4.

a. $A = 3.49e^{0.0292*10}$. Or, `=3.49*EXP(0.0292*10)`. The cost of gas would be $4.67.

b. $A = 1.99e^{0.0292*10}$. Or, `=1.99*EXP(0.0292*10)`. The cost of eggs would be $2.66.

c. $A = 3.29e^{0.0292*10}$. Or, `=3.29*EXP(0.0292*10)`. The cost of bread would be $4.41.

d. $A = 79.99e^{0.0292*10}$. Or, `=79.99*EXP(0.0292*10)` The cell phone bill would be $107.11.

e. $A = 1.99e^{0.0292*10}$. Or, `=1.99*EXP(0.0292*10)` The cost of a song download would be $2.66.

2.6.5.

$P = 3500 - 350 = \$3150$

$I = 3150(0.13)(2.33) \approx \$955.50$

You would pay $4,105.50 in 28 months with $146.63 as your monthly payments.

2.6.6.

$P(0.096)(5) + P = 15253.8$

The amount borrowed was $10,306.62.

2.6.7.

a. $A = \frac{350((1+\frac{0.065}{12})^{12*25}-1)}{\frac{0.065}{12}}$

Or, `=FV(0.065/12, 12*25, 350, 0)`

The future value is $262,092.78 and the interest earned is $157,092.78.

b. $A = \frac{500((1+\frac{0.065}{4})^{4*15}-1)}{\frac{0.065}{4}}$

Or, `=FV(0.065/4, 4*15, 500, 0)`

The future value is $50,168.34 and the interest earned is $20,168.34.

c. $A = \frac{75((1+\frac{0.045}{52})^{52*30}-1)}{\frac{0.045}{52}}$

Or, `=FV(0.045/52, 52*30, 75, 0)`

The future value is $247,448.43 and the interest earned is $130,448.43.

2.6.8.

House 1: $D = \frac{239920\frac{0.05}{12}}{(1-(1+\frac{0.05}{12}^{-12*30}))}$

```
=PMT(0.05/12, 12*30, 239920, 0)
```

Monthly payment \$9,996.67, Total Paid \$463,658.40, Interest Paid \$223,738.40

House 2: $D = \frac{311920\frac{0.05}{12}}{(1-(1+\frac{0.05}{12}^{-12*30}))}$

```
=PMT(0.05/12, 12*30, 3119200, 0)
```

Monthly Payment \$1,674.45, Total Paid \$602,803.45, Interest Paid \$290,883.45

2.6.9.

The net monthly cash flow is \$428.

2.6.10.

Pat is spending \$1,371 a year painting.

2.6.11.

$1000(0.12) = 120$. Your tax bill will be decreased by \$120.

2.6.12.

Your tax bill will be decreased by \$1,000.

2.6.13.

a. $43,000 + 1,000 = 44,000$. Amir's gross income is \$44,000.

b. $44,000 - 3,000 = 41000$. Amir's adjusted gross income is \$41,000.

2.6.14.

a. $41,000 - 12,000 = 29,000$. Amir's taxable income is \$29,000.

b.

$$\begin{aligned} &= 0.1(9525) + 0.12(29000 - 9525) \\ &= 952.5 + 2337 \\ &= 3289.5 \end{aligned}$$

Amir owes \$3289.50 in taxes.

2.6.15.

a. $3289.5 - 1200 - 2700 = -610.5$. Amir will be refunded \$610.50.

3 · Statistics

3.1 · Overview of the Statistical Process

3.1.22 · Exercises

3.1.22.1.

A sample is a sub group of the population. A population is the entire group of subjects.

3.1.22.2.

A Statistic is a measurement obtained from the data taken from a sample. A parameter is a measurement obtained from the data of the entire population.

3.1.22.3.

The intended population is all PCC students.

a. All PCC Students

b. 200 students from PCC Cascade campus

c. The collected data is not representative of all PCC students since it only includes responses from students at Cascade campus. This is an example of sampling bias.

3.1.22.4.

a. The intended population is all Washington County residences.

b. 1200 homes in Washington County

c. The collected data is likely representative since residences were selected at random from the entire county.

3.1.22.5.

a. The representatives in a state's congress.

b. The population size is $n = 106$

c. The sample size is in $n = 28$

d. The statistic is $\frac{14}{28} = 0.5$ or 50%

e. The confidence interval is $(45\%, 55\%)$ and tells us that the true percentage of the state congress representatives in support of the new education (the parameter) likely lies between 45% and 55%.

3.1.22.6.

a. All registered voters in the city of Raleigh.

b. The population size is $n = 9500$

c. The sample size is $n = 350$

d. The statistic is $\frac{112}{350} = 0.32$ or 32%

e. The confidence interval is $(28.5\%, 35.5\%)$ and tells us that the true percentage of registered voters who will vote for Brown is likely to lie between 28.5% and 35.5%.

3.1.22.7.

The population is all trout in the lake. The sample is the 20 that were caught.

3.1.22.8.

The population is all trees in the park. The sample is the 45 that were tagged and measured.

3.1.22.9.

Parameter

3.1.22.10.

A sample was taken so it is a statistic.

3.1.22.11.

Statistic

3.1.22.12.

a. Stratified
b. Simple Random Sample
c. Systematic

3.1.22.13.

a. Stratified
b. Volunteer
c. Simple Random Sample

3.1.22.14.

a. Volunteer Bias
b. Sampling Bias
c. Response Bias
d. Non Response Bias
e. Response Bias
f. Loaded Question

3.1.22.15.

a. Loaded Question
b. Volunteer Bias
c. Response Bias
d. Volunteer
e. Response Bias
f. Response Vias or Non-response Bias

3.1.22.16.

a. Observational study
b. Experiment
c. Observational study

3.1.22.17.

a. Observational study

b. Experiment

c. Observational study

3.1.22.18.

a. Group 1

b. Group 2

c. Blind because the patients in the study do not know.

d. Controlled experiment

3.1.22.19.

a. Cancer patients

b. No because sampling has variability

c. Stratified

d. Convenient Sample. It does not represent the population.

3.1.22.20.

a. 2nd Group

b. Inert pill group

c. Double Blind because the patients and the advisors do not know who is in each group.

d. Placebo-controlled experiment

3.1.22.21.

a. All students

b. Experiment

c. It is only looking at one class and not all groups that are in the population so Subjects are not randomly sampled from a specified population.

3.1.22.22.

The control group would be the group that were ask to tell the truth. It is a blind study because the person who is administering the lie detector test doesn't know which group each person is in.

3.1.22.23.

a. 0.05 or 5%

b. (25%, 35%)

c. I am confident that the percentage of college freshmen who prefer morning classes is between 25% to 35%.

3.1.22.24.

a. 3.5% or 0.035

b. (34.5%, 41.5%)

c. I am confident that the percentage of all U.S. Employees are engaged at work is between 34.5% and 41.5%.

3.1.22.25.

a. $(24+36)/2=30$. The statistic is 30%.

b. $30-24=6$. The margin of error is 6%.

3.1.22.26.

a. $(44+52)/2=48$. The statistic is 48%.

b. $48-44=4$. The margin of error is 4%.

3.1.22.27.

Play Barry Manilow to half the crop and don't play any music to the other half of the crop.

3.1.22.28.

Answer will vary. Here is one potential solution.

Observational: Survey adults who played sports as a child. Ask them if they consider themselves extroverted or introverted. Record the percentage of each personality type.

Experimental: Survey adults and ask them if they played sports as children. Ask them if they consider themselves extroverted or introverted. Compare the results of adults who played sports as children with the adults who didn't play sports as children.

Which is more practical? The Experimental study would allow you to see if there is a significant difference in the personality types based on if the person played sports or didn't play sports.

3.1.22.29.

Answers will vary depending on the article.

3.1.22.30.

Answers will vary depending on the poll being studied.

3.2 · Describing Data

3.2.24 · Exercises

3.2.24.1.

True

3.2.24.2.

False

3.2.24.3.

a. Quantitative

b. Categorical

c. Categorical

d. Quantitative
e. Quantitative

3.2.24.4.

a. Quantitative
b. Categorical
c. Quantitative
d. Quantitative
e. Quantitative

3.2.24.5.

Bar graphs and pie charts are used for categorical data.

3.2.24.6.

Histograms are used for quantitative data.

3.2.24.7.

a. 2 had 3 children.
b. 15 adults were questioned.
c. 33.33% of the adults questioned had 0 children.

3.2.24.8.

a. 8 movies took 2 days to arrive.
b. She ordered 19 movies total.
c. 21.05% of the movies arrived in one day.

3.2.24.9.

5 students earned and A on their paper.

3.2.24.10.

24 served drinks were lattes.

3.2.24.11.

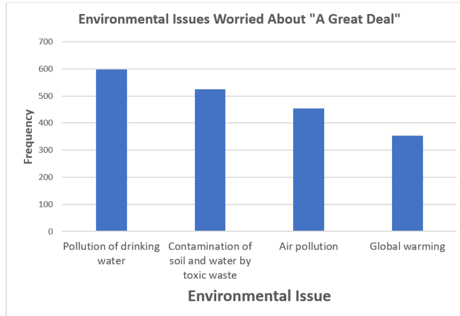
Cory spent $676 on rent this month.

3.2.24.12.

Habiba spends 11.8 hours travelling each week.

3.2.24.13.

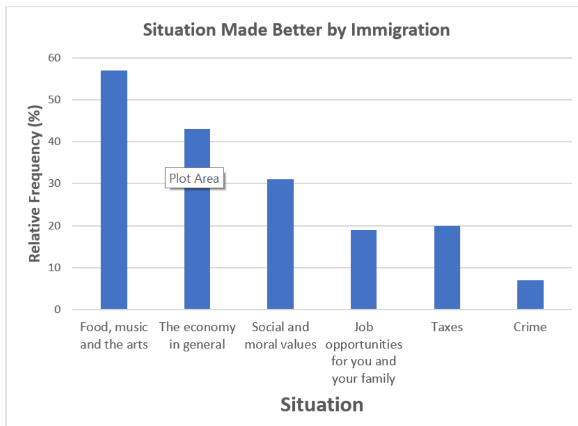
a. This data is categorical.

b. 

c. No, we cannot make a pie chart out of this data. The total of the relative frequencies is 1932, but only 1012 adults were asked. So some adults selected multiple options.

3.2.24.14.

a. This data is categorical.

b. 

c. No, we cannot make a pie chart out of this data. The relative frequencies add to 177%, which is more than 100%.

3.2.24.15.

a. 40 households heat their home with firewood.

b. 50% of the households heat their home with natural gas.

3.2.24.16.

a. 54% of the students are below senior class.

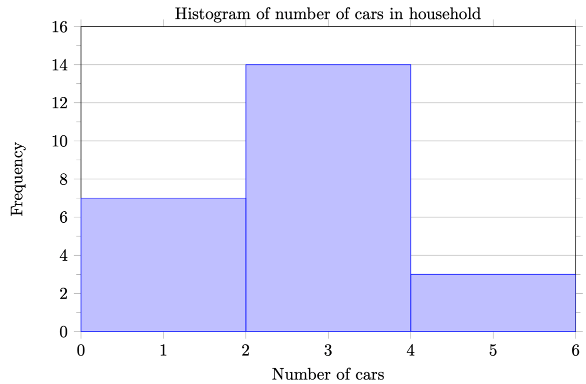
b. 9 of the sampled students are freshmen.

3.2.24.17.

a. This data is qualitative

b.

Number of cars in household	Frequency
0-1	7
2-3	14
4-5	3

c. 

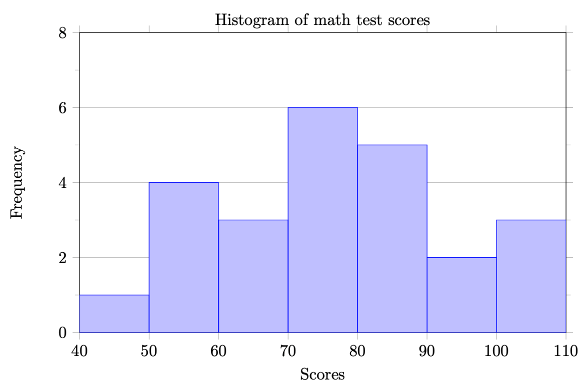
d. This data is unimodal and skewed right, with no outliers.

3.2.24.18.

a. This is quantitative data.

b.

Math Test Score	Frequency
40-49	1
50-59	4
60-69	3
70-79	6
80-89	5
90-99	2
100-109	3

c. 

d. The data is unimodal. There are no obvious outliers. The data is either symmetric or skewed left.

3.2.24.19.

The graph would be more effective at displaying the true differences between the categories if the vertical scale started at 0. The vertical axis is missing a label and units, so we can't tell if those are frequencies or relative frequencies. A flat bar graph (instead of the 3d graph) would be easier to read.

3.2.24.20.

No, this chart does not present a good representation of this data. The percentages in a pie chart must add to 100%, but these add to 193%. A bar chart would be appropriate for this data.

3.2.24.21.

This is a poor graph because the vertical axis does not have a numerical scale, so we cannot know how many have each drink as a favorite drink. We also don't know if the bottom vertical line represents 0, which is potentially misleading.

3.2.24.22.

2This is a misleading graph because the $-1320 in April represents a loss (instead of a profit), but the height of the bar for April looks like a profit. Also, the horizontal axis label is "Year", but the bar labels are months.

3.2.24.23.

a. Normal distribution – The number of heads in 24 sets of 100 coin flips.

b. Positive or right skewed – Distribution of scores on a psychology test.

c. Negative or left skewed – Scores on a 20-point statistics quiz.

d. Bimodal – The frequency of times between eruptions of the Old Faithful geyser.

3.2.24.24.

a. This distribution is unimodal and right skewed. There is a possible outlier between 10 and 11.

b. This distribution is unimodal and symmetric. There are no outliers.

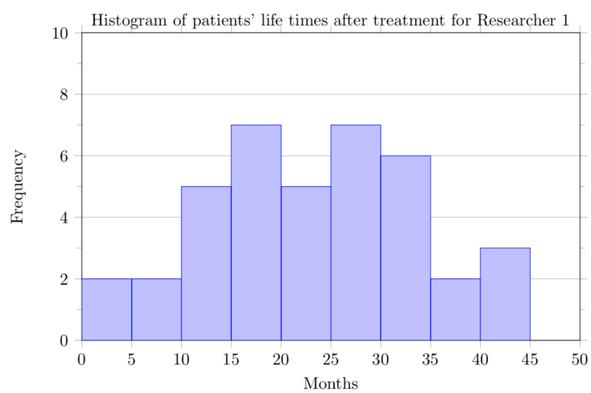
c. This distribution is unimodal and right skewed. There are no obvious outliers.

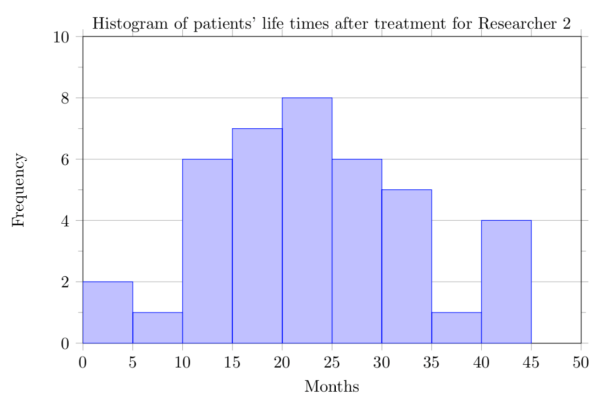
d. This distribution is multimodal and left skewed. There are no outliers.

e. This distribution is unimodal and left skewed. There are no obvious outliers.

3.2.24.25.

a.





b. The data for patients of both researchers is symmetric. Researcher's 1 patients' data appears to be unimodal, but Researcher 2's patients' data may be bimodal or multimodal. The data for Researcher 1's patients does not have any outliers, but the data for Researcher's 2 may have outliers between 0 and

5 months or 40 and 45 months.

3.3 · Summary Statistics: Measures of Center

3.3.7 · Exercises

3.3.7.1.

a. In Excel:

`=average(7.50,25.00,10.00,10.00,7.50,8.25,9.00,5.00,15.00,8.00,7.25,7.50,8.00,7.00,12.00)`

= \$9.80

There are 15 times shown, so $n = 15$. The mean is:

$$\bar{x} = \frac{(7.50 + 25.00 + 10.00 + 10.00 + 7.50 + 8.25 + 9.00 + 5.00 + 15.00 + 8.00 + 7.25 + 7.50 + 8.00 + 7.00 + 12.00)}{15}$$
$$= \$9.80$$

b. In Excel:

`=median(7.50,25.00,10.00,10.00,7.50,8.25,9.00,5.00,15.00,8.00,7.25,7.50,8.00,7.00,12.00)`

= \$8.00

There are 15 times shown, so $n = 15$. We start by listing the data in order:

\$5.00, \$7.00, \$7.25, \$7.50, \$7.50, \$7.50, \$8.00, \$8.00, \$8.25, \$9.00, \$10.00, \$10.00, \$12.00, \$15.00, \$25.00

$Median = \$8.00$

c. Since the mean is greater than the median, we would expect the distribution will be skewed right.

3.3.7.2.

a. In Excel:

`=average(10, 12.75, 7, 9, 9.75, 6.5, 12.5, 12.5, 8.75, 17, 10.5, 2)`

= 9.85 minutes

There are 12 times shown, so $n = 12$. The mean is:

$$\bar{x} = \frac{(10 + 12.75 + 7 + 9 + 9.75 + 6.5 + 12.5 + 12.5 + 8.75 + 17 + 10.5 + 2)}{12}$$
$$= 9.85 \text{ minutes}$$

b. In Excel:

`=median(10, 12.75, 7, 9, 9.75, 6.5, 12.5, 12.5, 8.75, 17, 10.5, 2)`

= 9.88 minutes

There are 12 times shown, so $n = 12$. We start by listing the data in order:

2, 6.5, 7, 8.75, 9, 9.75, 10, 10.5, 12.5, 12.5, 12.75, 17

$$Median = \frac{9.75 + 10}{2}$$
$$= 9.88 \text{ minutes}$$

c. Because the mean and median are approximately equal, we would expect that the distribution is symmetric.

3.3.7.3.

a. In Excel:

`=average(15.2,18.8,19.3,19.7,20.2,21.8,22.1,29.4)`

$= 20.81$ seconds

There are 8 times shown, so $n = 8$.

$$\bar{x} = \frac{(15.2 + 18.8 + 19.3 + 19.7 + 20.2 + 21.8 + 22.1 + 29.4)}{15}$$
$$= 20.81 \text{ seconds}$$

b. In Excel:

`=median(15.2,18.8,19.3,19.7,20.2,21.8,22.1,29.4)`

$= 19.95$ seconds

There are 8 times shown, so $n = 8$. The times are given already in order:

15.2, 18.8, 19.3, 19.7, 20.2, 21.8, 22.1, 29.4

$$Median = \frac{19.7 + 20.2}{2}$$
$$= 19.95 \text{ seconds}$$

c. Since the mean and median are approximately equal, we would expect that the distribution is symmetric.

3.3.7.4.

a. In Excel:

`=average(3.49,3.51,3.51,3.51,3.52,3.54,3.55,3.58,3.61)`

$= 3.536$ grams

There are 9 weights shown, so $n = 9$.

$$\bar{x} = \frac{(3.49 + 3.51 + 3.51 + 3.51 + 3.52 + 3.54 + 3.55 + 3.58 + 3.61)}{9}$$
$$= 3.536 \text{ grams}$$

b. In Excel:

`=median(3.49, 3.51, 3.51, 3.51, 3.52, 3.54, 3.55, 3.58, 3.61)`

$= 3.52$ grams

There are 9 weights shown, so $n = 9$. The weights are already given in order:

3.49, 3.51, 3.51, 3.51, 3.52, 3.54, 3.55, 3.58, 3.61

$$Median = 3.52 \text{ grams}$$

c. Since the mean and median are approximately equal, we would expect the distribution to be symmetric.

3.3.7.5.

a. In GeoGebra Classic, enter the costs into the column A and frequencies into column B of the spreadsheet and use the “One Variable Analysis” function. Then use the “Show Statistics” option.

$$Mean = 33.8 \text{ thousand dollars}$$

The sum of the frequencies is 75, so $n = 75$

$$\begin{aligned}\bar{x} &= \frac{153 + 207 + 2510 + 3015 + 3513 + 4011 + 459 + 507}{75} \\ &= 33.8 \text{ thousand dollars}\end{aligned}$$

b. In GeoGebra Classic, enter the costs into the column A and frequencies into column B of the spreadsheet and use the “One Variable Analysis” function. Then use the “Show Statistics” option.

$$Median = 35 \text{ thousand dollars}$$

Since there are 75 values (an odd number), we know that the median will be the single middle data value. Because $\frac{75}{2} = 37.5$, we know it will be the 38th value in the list. The 38th value is 35, so the median is 35 thousand dollars.

c. Since the mean is less than the median, we would expect the distribution to be skewed left.

3.3.7.6.

a. In GeoGebra Classic, enter the costs into the column A and frequencies into column B of the spreadsheet and use the “One Variable Analysis” function. Then use the “Show Statistics” option.

$$Mean = 5.3 \text{ thousand characters}$$

The sum of the frequencies is 34, so $n = 34$

$$\begin{aligned}\bar{x} &= \frac{(04 + 15 + 22 + 33 + 43 + 51 + 63 + 73 + 93 + 103 + 112 + 142)}{34} \\ &= 5.3 \text{ thousand characters}\end{aligned}$$

b. In GeoGebra Classic, enter the costs into the column A and frequencies into column B of the spreadsheet and use the “One Variable Analysis” function. Then use the “Show Statistics” option.

$$Median = 4.5 \text{ thousand characters}$$

Since there are 34 values (an even number), we know that the median will be the mean of the two middle values. Because $\frac{34}{2} = 17$, we know the two middle values are the 17th and 18th values. The 17th value is 4, and the 18th value is 5, so the median is 4.5 thousand characters.

c. Since the mean is greater than the median, we expect the distribution to be skewed right.

3.3.7.7.

a. For Researcher 1:

   In Excel:

   `=average(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`

   = 23.6 months.

   `=median(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,`

   = 24 months

   The mean for Researcher 1's patients is 23.6 months, and the median for Researcher's 1 patients is 24 months.

   For Researcher 2:

   In Excel:

   `=average(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`

   = 22.8 months

   `=median(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,`

   = 22 months

   The mean for Researcher 2's patients is 22.8 months, and the median is 22 months

b. Both the mean and median for Researcher 1's patients are greater than the mean and median for Researcher 2's patients. So, on average, Researcher 1's patients have a longer life time after starting the cancer treatment than Researcher 2's patients.

3.3.7.8.

a. For Males:

   In Excel

   ```
   =average(53000,70000,12800,30000,4500,42000,48000,60000,108000,11000)
   ```

   $= \$43,930$

   ```
   =median(53000,70000,12800,30000,4500,42000,48000,60000,108000,11000)
   ```

   $= \$45,000$

   The mean for males is $43,930, and the median for males is $45,000.

   For Females:

   ```
   =average(1600,1200,20000,25000,670,29000,44000,30000,5800,50000)
   ```

   $= \$20,727$

   ```
   =median(1600,1200,20000,25000,670,29000,44000,30000,5800,50000)
   ```

   $= \$22,500$

   The mean for females is $20,727, and the median for females is $22,500

b. Both the mean income and median income for the males in the sample were twice as large as the mean income and median income for the females in the sample. The difference between the mean and median income for the females was slightly larger than the mean and median income for the males, so the distribution of incomes for the females is possibly more skewed than the distribution for the males.

3.3.7.9.

GeoGebra was used to create the histograms. You should check with your instructor to see if histograms are to be hand-drawn or computer generated. Answers will vary depending on the size of the margins and the programs you are using.

a.

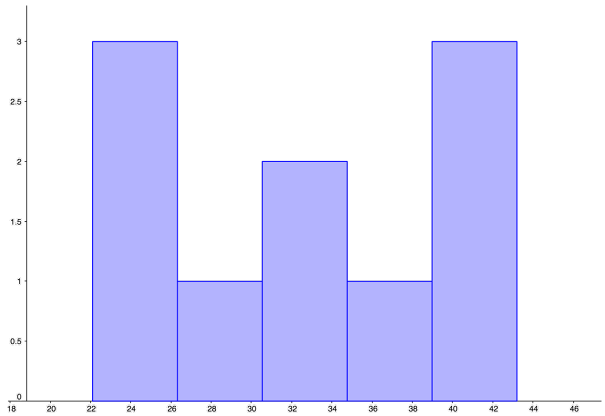


Figure .0.7 Histogram for Average Number of Pieces Correctly Remembered by Non-players

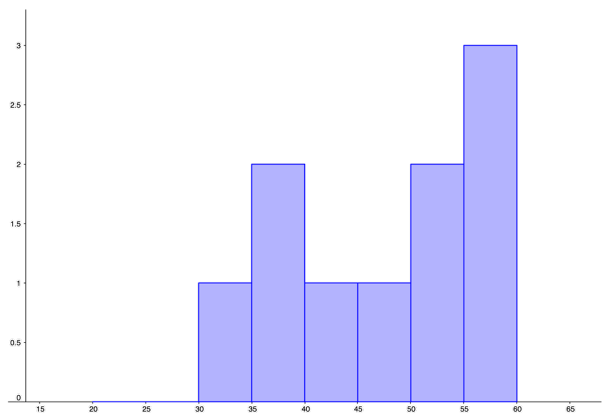


Figure .0.8 Histogram for Average Number of Pieces Correctly Remembered by Beginners

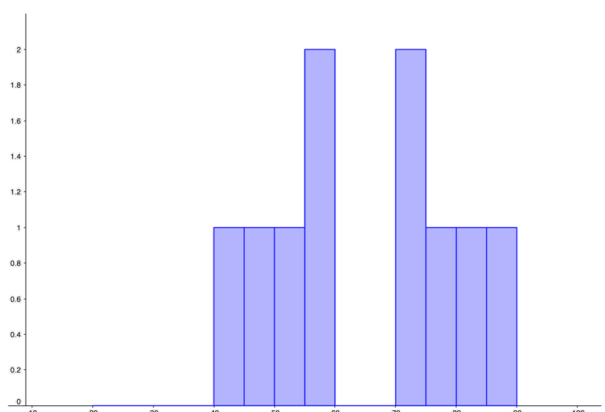


Figure .0.9 Histogram for Average Number of Pieces Correctly Remembered by Tournament Players

b. The mean number of pieces correctly remembered for non-players was 33.65 pieces.

The mean number of pieces correctly remembered for beginners was 47.6 pieces.

The mean number of pieces correctly remembered for tournament players was 64.98 pieces.

c. The median number of pieces correctly remembered for non-players was 33.5 pieces.

   The median number of pieces correctly remembered for beginners was 51.3 pieces.

   The median number of pieces correctly remembered for tournament players was 71.1 pieces.

d. The distribution for non-players appears to be uniform. The distribution for beginners looks unimodal and left-skewed. The distribution for tournament players appears bimodal and symmetric.

   The mean and median number of pieces correctly remembered were both greatest for tournament players, with non-players having the smallest mean and median of pieces correctly remembered.

3.3.7.10.

a.

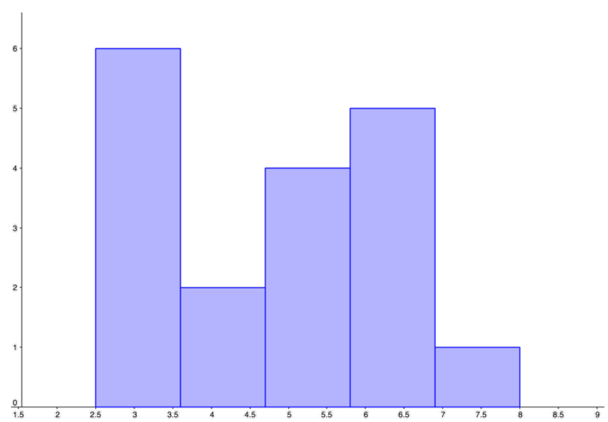


Figure .0.10 Histogram for Smile-Leniency of False Smiles

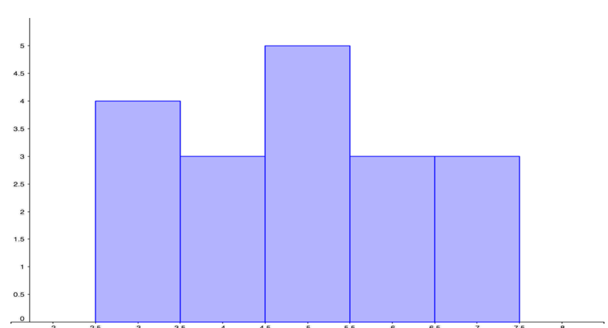


Figure .0.11 Histogram for Smile-Leniency of Felt Smiless

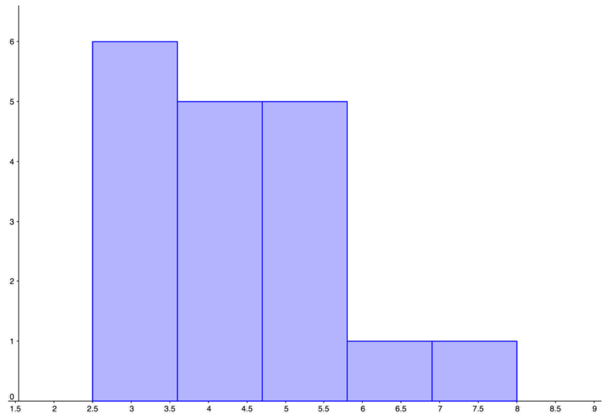


Figure .0.12 Histogram for Smile-Leniency of Miserable Smiles

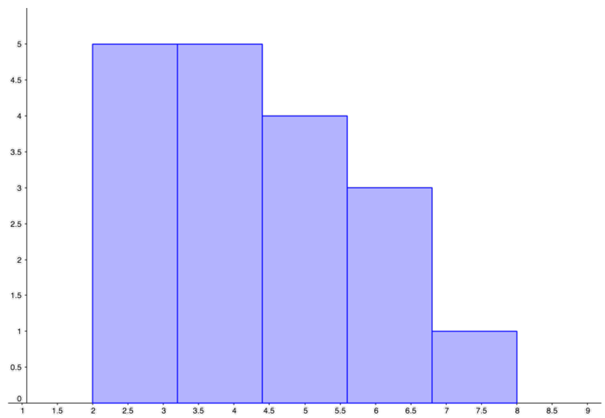


Figure .0.13 Histogram for Smile-Leniency of Neutral Control Group

b. The mean leniency for the false smile group was 5.4.

   The mean leniency for the felt smile group was 4.9.

   The mean leniency for the miserable smile group was 4.9.

   The mean leniency for the neutral control group was 4.1.

c. The median leniency for the false smile group was 5.5.

   The median leniency for the felt smile group was 4.8.

   The median leniency for the miserable smile group was 4.8.

   The median leniency for the neutral control group was 4.0.

d. Answers will vary depending on the graphs created. The shape of the false smile is bimodal, whereas the miserable smile and neutral control groups are both unimodal. The miserable smile and neutral control both appear to be skewed to the right, however, comparing the mean to the median we see that there is not much of a difference. Therefore, none of the graphs are skewed. The felt smile group is the most visually uniform. The measures of center (mean and median) have the center of the false smile near 5.5, the felt smile and miserable smile both near 4.8, and the control near 4

3.3.7.11.

a. There are many possible answers for this problem. Three data sets with 5 values each that have the same mean but different medians are:

$$\begin{array}{lllll} 0, & 0, & 0, & 0, & 10 \\ 0, & 0, & 2, & 4, & 4 \\ 0, & 1, & 1, & 1, & 7 \end{array}$$

b. There are many possible answers for this problem. Three data sets with 5 values that have the same median but different means are:

$$\begin{array}{lllll} 10, & 10, & 10, & 10, & 10 \\ 0, & 0, & 10, & 15, & 20 \\ 1, & 5, & 10, & 10, & 10 \end{array}$$

3.3.7.12.

a. Argument for categorical: Because it is not clear that the shoe sizes represent a measure, the data can be considered categorical.

Argument for quantitative: Because shoe size is a measurement that corresponds to the length of someone's foot, it can be treated as quantitative data.

b. Each graph would have frequency along the y-axis. In a bar graph the bars would have spaces between them and each bar would be labeled with the shoe size. In a histogram there would not be any spaces between the bars and the shoe sizes could be the scale on the x-axis.

c. The mean shoe size to be 7.2 and the median shoe size to be 7.

The mean shoe size is:

$$\begin{aligned} \bar{x} &= \frac{(54 + 64 + 76 + 86 + 95)}{25} \\ &= 7.2 \end{aligned}$$

Since there are 25 values (an odd number), we know that the median will be the single middle data value. Because $\frac{25}{2} = 12.5$, we know it will be the 13th value in the list. The 13th value is 7, so the median shoe size is 7.

The mean and median shoe size might be useful statistics to the store. If shoe size is positively correlated to height, then a shoe store with a comparatively larger mean or median shoe size could determine that their clients are, on average, comparatively taller. (Other answers are possible.)

3.3.7.13.

a. This graph is skewed left.

b. I expect that the mean is less than the median because the graph is skewed left.

3.3.7.14.

a. This graph is symmetric. The graph has a single peak between $65,000 an $66,000, and there are approximately an equal number of data values on either side of this peak.

b. I expect that the mean and median are equal because the graph is symmetric.

3.3.7.15.

Answers will vary.

3.4 · Summary Statistics: Measures of Variation

3.4.10 · Exercises

3.4.10.1.

a. In Excel:

Entering the data values into cells A1 through A15.

$$\begin{aligned} s &= stdev.s(A1 : A15) \\ &= \$4.82 \end{aligned}$$

From Exercise 3.3.7.1, the mean is \$9.80. There are 15 data values, so $n = 15$.

We will make a table of data values, their deviations from the mean, and the squared deviations:

Data Value	Deviation	Deviation Squared
7.5	$7.5 - 9.8 = -2.3$	$(-2.3)^2 = 5.29$
25	$25 - 9.8 = 15.2$	$(15.2)^2 = 231.04$
10	$10 - 9.8 = 0.2$	$(0.2)^2 = 0.04$
10	$10 - 9.8 = 0.2$	$(0.2)^2 = 0.04$
7.5	$7.5 - 9.8 = -2.3$	$(-2.3)^2 = 5.29$
8.25	$8.25 - 9.8 = -1.55$	$(-1.55)^2 = 2.4$
9	$9 - 9.8 = -0.8$	$(-0.8)^2 = 0.64$
5	$5 - 9.8 = -4.8$	$(-4.8)^2 = 23.04$
15	$15 - 9.8 = 5.2$	$(5.2)^2 = 27.04$
8	$8 - 9.8 = -1.8$	$(-1.8)^2 = 3.24$
7.25	$7.25 - 9.8 = -2.55$	$(-2.55)^2 = 6.5$
7.5	$7.5 - 9.8 = -2.3$	$(-2.3)^2 = 5.29$
8	$8 - 9.8 = -1.8$	$(-1.8)^2 = 3.24$
7	$7 - 9.8 = -2.8$	$(-2.8)^2 = 7.84$
12	$12 - 9.8 = 2.2$	$(2.2)^2 = 4.84$

Next, we add the squared deviations and get $5.29 + 231.04 + 0.04 + 0.04 + 5.29 + 2.4 + 0.64 + 23.04 + 27.04 + 3.24 + 6.5 + 5.29 + 3.24 + 7.84 + 4.84 = 325.78$ dollars-squared.

The sample standard deviation is:

$$\begin{aligned} s &= \sqrt{\frac{325.78}{14}} \\ &= \$4.82 \end{aligned}$$

b. In GeoGebra:

In GeoGebra Classic, enter the data values into the column A of the spreadsheet and use the “One Variable Analysis” function. Then use the “Show Statistics” option to find the five-number summary.

Min	Q1	Median	Q3	Max
\$5	\$7.50	\$8	\$10	\$25

From Exercise 3.3.7.1, the data listed in order is:

\$5.00, \$7.00, \$7.25, \$7.50, \$7.50, \$7.50, \$8.00, \$8.00, \$8.25, \$9.00, \$10.00, \$10.00, \$12.00, \$15.00, \$25.00

Also from Exercise 3.3.7.1, there are 15 data values ($n = 15$), and the median is \$8.00. The lower half of the data is:

\$5.00, \$7.00, \$7.25, \$7.50, \$7.50, \$7.50, \$8.00

The median of the lower half is \$7.50, so the lower quartile Q_1 is \$7.50.

The upper half of the data is:

\$8.25, \$9.00, \$10.00, \$10.00, \$12.00, \$15.00, \$25.00

The median of the upper half is \$10.00, so the upper quartile Q_3 is \$10.00.

The smallest and largest data values are \$5.00 and 25.00, respectively, so the min and max are \$5.00 and \$25.00. The five-number summary is:

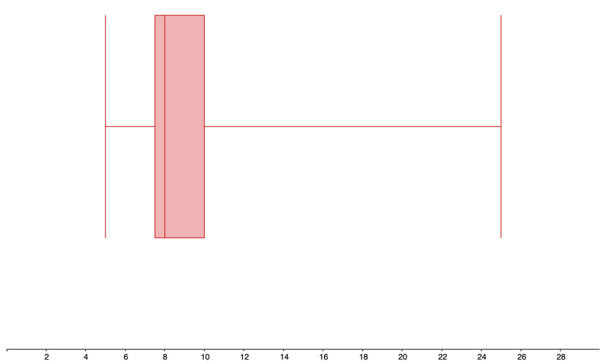
Min	Q1	Median	Q3	Max
\$5	\$7.50	\$8	\$10	\$25

c. The range is:

$$\begin{aligned} Range &= Max - Min \\ &= 25 - 5 \\ &= \$20 \end{aligned}$$

The interquartile range (IQR) is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 10 - 7.5 \\ &= \$2.50 \end{aligned}$$

d. 

3.4.10.2.

a. In Excel:

I entered the data values into cells A1 through A12.

The standard deviation is:

$$\begin{aligned} s &= stdev.s(A1 : A12) \\ &= 3.78 \text{ hours} \end{aligned}$$

b. In GeoGebra:

In GeoGebra Classic, enter the data values into the column A of the spreadsheet and use the "One Variable Analysis" function. Then use the "Show Statistics" option to find the five-number summary.

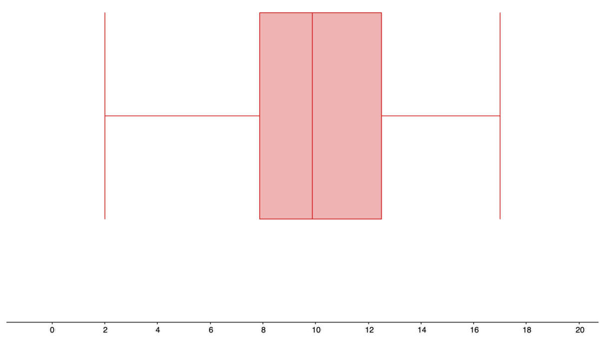
Min	Q1	Median	Q3	Max
2 hours	7.875 hours	9.875 hours	12.5 hours	17 hours

c. The range is:

$$\begin{aligned} Range &= Max - Min \\ &= 17 - 2 \\ &= 15 \text{ hours} \end{aligned}$$

The interquartile range (IQR) is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 12.5 - 7.875 \\ &= 3.625 \text{ hours} \end{aligned}$$

d. 

3.4.10.3.

a. In Excel:

I entered the data values into cells A1 through A9.

The standard deviation is:

$$\begin{aligned} s &= stdev.s(A1 : A9) \\ &= 4.068 \text{ seconds} \end{aligned}$$

b. In GeoGebra:

In GeoGebra Classic, enter the data values into the column A of the spreadsheet and use the “One Variable Analysis” function. Then use the “Show Statistics” option to find the five-number summary.

Min	Q1	Median	Q3	Max
15.2 seconds	19.05 seconds	19.95 seconds	21.95 seconds	29.4 seconds

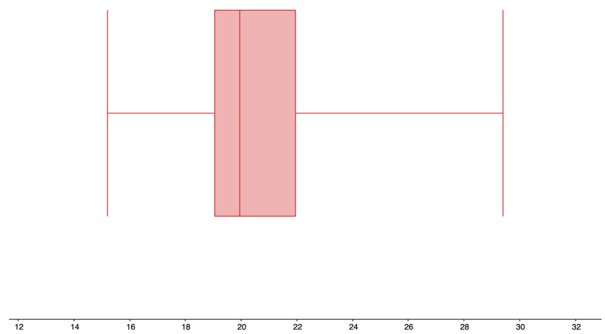
c. The range is:

$$Range = Max - Min$$

$$= 29.4 - 15.2$$
$$= 14.2 \text{ seconds}$$

The interquartile range (IQR) is:

$$IQR = Q_3 - Q_1$$
$$= 21.95 - 19.05$$
$$= 2.9 \text{ seconds}$$

d. 

3.4.10.4.

a. In Excel:

I entered the data values into cells A1 through A9.

The standard deviation is:

$$s = stdev.s(A1 : A9)$$
$$= 0.039 \text{ grams}$$

b. In GeoGebra:

In GeoGebra Classic, enter the data values into the column A of the spreadsheet and use the "One Variable Analysis" function. Then use the "Show Statistics" option to find the five-number summary.

Min	Q1	Median	Q3	Max
3.49 grams	3.51 grams	3.52 grams	3.565 grams	3.61 grams

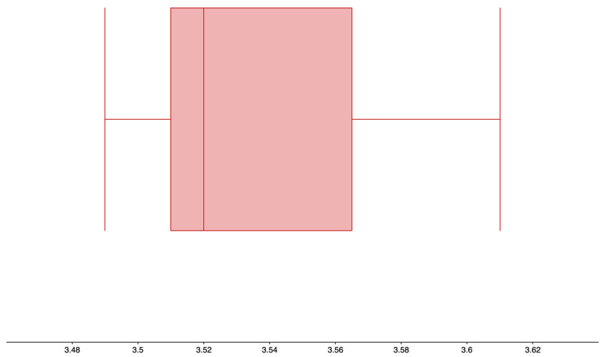
c. The range is:

$$Range = Max - Min$$
$$= 3.61 - 3.49$$
$$= 0.12 \text{ grams}$$

The interquartile range (IQR) is:

$$IQR = Q_3 - Q_1$$

$$= 3.565 - 3.51$$
$$= 0.055 \text{ grams}$$

d. 

3.4.10.5.

a. In GeoGebra:

In GeoGebra Classic, enter the costs into column A and frequencies into column B of the spreadsheet and use the "One Variable Analysis" function. Then use the "Show Statistics" option. The standard deviation is:

$$s = 9.58 \text{ thousand dollars}$$

From Exercise 3.3.7.5, the mean is 33.8 thousand dollars.

The mean and the standaed deviation together tell us that, on average, the cars at the local dealership are $9,580 from the mean price of $33,800.

b. In GeoGebra:

In GeoGebra Classic, enter the data values into the column A of the spreadsheet and use the "One Variable Analysis" function. Then use the "Show Statistics" option to find the five-number summary.

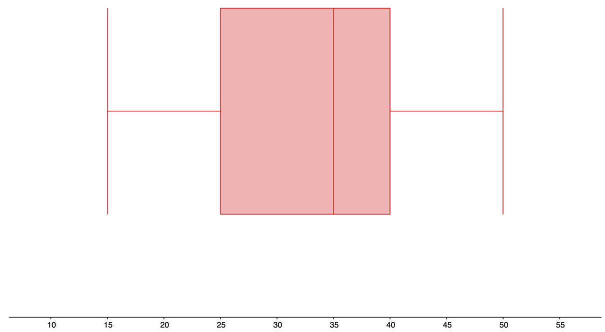
Min	Q1	Median	Q3	Max
15 thousand dollars	25 thousand dollars	35 thousand dollars	40 thousand dollars	50 thousand dollars

c. The range is:

$$\begin{aligned} Range &= Max - Min \\ &= 50 - 15 \\ &= 35 \text{ thousand dollars} \end{aligned}$$

The interquartile range (IQR) is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 40 - 15 \\ &= 25 \text{ thousand dollars} \end{aligned}$$

d. 

3.4.10.6.

a. In GeoGebra:

In GeoGebra Classic, enter the costs into column A and frequencies into column B of the spreadsheet and use the "One Variable Analysis" function. Then use the "Show Statistics" option. The standard deviation is:

$$s = 4.20 \text{ thousand characters}$$

From Exercise 3.3.7.6, the mean is 5.3 thousand characters.

The mean and standard deviation together tell us that on average the emails vary from the mean of 5.3 thousand characters by 4.2 thousand characters.

b. In GeoGebra:

In GeoGebra Classic, enter the data values into the column A of the spreadsheet and use the "One Variable Analysis" function. Then use the "Show Statistics" option to find the five-number summary.

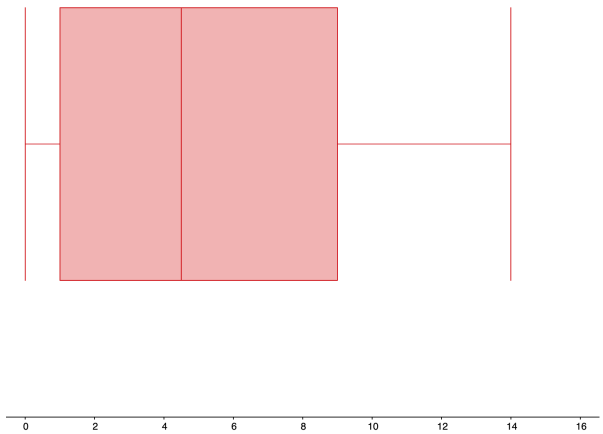
Min	Q1	Median	Q3	Max
0 thousand characters	1 thousand characters	4.5 thousand characters	9 thousand characters	14 thousand characters

c. The range is:

$$\begin{aligned} Range &= Max - Min \\ &= 14 - 0 \\ &= 14 \text{ thousand characters} \end{aligned}$$

The interquartile range (IQR) is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 9 - 1 \\ &= 8 \text{ thousand characters} \end{aligned}$$

d. 

3.4.10.7.

a. In GeoGebra:

   In GeoGebra Classic, enter the data values for Researcher 1 into column A of the spreadsheet, and enter the data values for Researcher 2 into column B of the spreadsheet. Then use the “Multiple Variable Analysis” function. Then use the “Show Statistics” function to display the sample standard deviation for each set of data values.

   The sample standard deviation for Researcher 1 is 11.25 months. The sample standard deviation for Research 2 is 11.38 months.

b. In GeoGebra:

   In GeoGebra Classic, enter the data values for Researcher 1 into column A of the spreadsheet, and enter the data values for Researcher 2 into column B of the spreadsheet. Then use the “Multiple Variable Analysis” function. Then use the “Show Statistics” function to display the sample standard deviation for each set of data values.

   The 5-number summary for Researcher 1 is:

Min	Q1	Median	Q3	Max
3 months	15 months	24 months	32.5 months	47 months

   The 5-number summary for Researcher 2 is:

Min	Q1	Median	Q3	Max
2 months	16 months	22 months	30 months	44 months

c. The range for Researcher 1 is:

$$\begin{aligned} Range &= Max - Min \\ &= 47 - 3 \\ &= 44 \text{ months} \end{aligned}$$

   The interquartile range (IQR) for researcher 1 is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 32.5 - 15 \\ &= 17.5 \text{ months} \end{aligned}$$

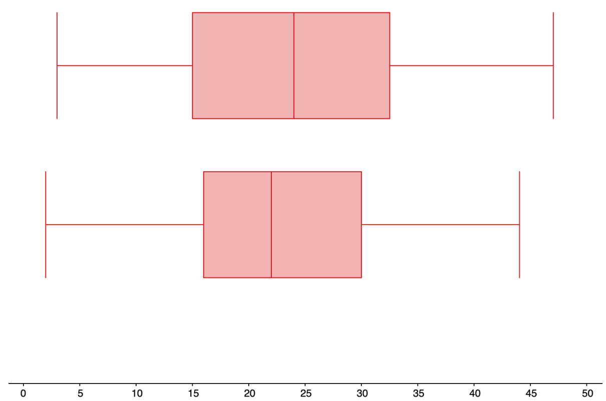
The range for Researcher 2 is:

$$\begin{aligned} Range &= Max - Min \\ &= 44 - 2 \\ &= 42 \text{ months} \end{aligned}$$

The interquartile range (IQR) for Researcher 2 is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 30 - 16 \\ &= 14 \text{ months} \end{aligned}$$

d. In GeoGebra Classic, enter the data values for Researcher 1 into column A of the spreadsheet, and enter the data values for Research 2 into column B of the spreadsheet. Then use the "Multiple Variable Analysis" function. Then select "Stacked BoxPlots" from the drop-down menu.



Researcher 1 has a larger minimum, median, 3rd quartile, and maximum than Researcher 2. For Researcher 1, 50% of the patients live longer than 24 months after treatment, compared to 50% of patients living longer than 22 months after treatment for Researcher 1.

Researcher 2 has less variation in the life times than Researcher 1, with an IQR of 14 months for Researcher 2, compared to an IQR 16.5 months for Researcher 1.

3.4.10.8.

a. In GeoGebra:

In GeoGebra Classic, enter the data values for males into column A of the spreadsheet, and enter the data values for females into column B of the spreadsheet. Then use the "Multiple Variable Analysis" function. Then use the "Show Statistics" function to display the sample standard deviation for each set of data values.

The sample standard deviation for males is $31,530.66. The sample standard deviation for females $18,806.22.

b. In GeoGebra:

In GeoGebra Classic, enter the data values for males into column A of the spreadsheet, and enter the data values for females into column B of the spreadsheet. Then use the "Multiple Variable Analysis" function. Then use the "Show Statistics" function to display the sample standard deviation for each set of data values.

The 5-number summary for males is:

Min	Q1	Median	Q3	Max
\$4,500	\$12,800	\$45,000	\$60,000	\$108,000

The 5-number summary for females is:

Min	Q1	Median	Q3	Max
\$670	\$1,600	\$22,500	\$30,000	\$50,000

c. The range for males is:

$$\begin{aligned} Range &= Max - Min \\ &= 108000 - 4500 \\ &= \$103,500 \end{aligned}$$

The interquartile range (IQR) for males is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 60000 - 12800 \\ &= \$47,200 \end{aligned}$$

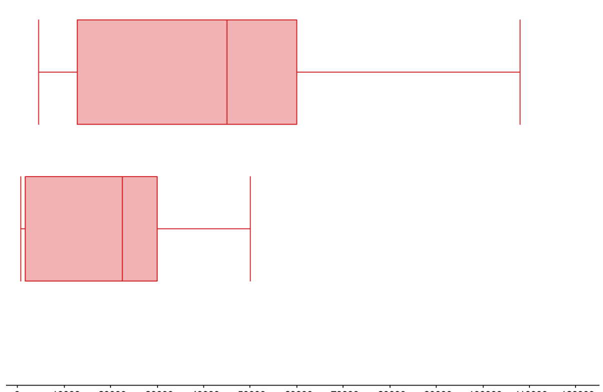
The range for females is:

$$\begin{aligned} Range &= Max - Min \\ &= 50000 - 670 \\ &= \$49,330 \end{aligned}$$

The interquartile range (IQR) for females is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 30000 - 1600 \\ &= \$28,400 \end{aligned}$$

d.



All of the values for the males' 5-number summary are larger than the corresponding values for the females' 5-number summary. The maximum income for males is more than twice the maximum income for females. Males have a significantly greater amount of variation in incomes than females, as shown by the larger range, IQR and standard deviation for males compared to females.

3.4.10.9.

a. In GeoGebra:

In GeoGebra Classic, enter the data values for non-players into column A of the spreadsheet, enter the data values for beginners into column B of the spreadsheet, and enter the data value for tournament players into column C of the spreadsheet. Then use the “Multiple Variable Analysis” function. Then use the “Show Statistics” function to display the sample standard deviation for each set of data values.

The sample standard deviation for non-players is 8.033 chess pieces. The sample standard deviation for beginners is 9.031 chess pieces. The sample standard deviation for tournament players is 15.622 chess pieces.

b. In GeoGebra:

In GeoGebra Classic, enter the data values for non-players into column A of the spreadsheet, enter the data values for beginners into column B of the spreadsheet, and enter the data value for tournament players into column C of the spreadsheet. Then use the “Multiple Variable Analysis” function. Then use the “Show Statistics” function to display the sample standard deviation for each set of data values.

The 5-number summary for non-players is:

Min	Q1	Median	Q3	Max
22.1 chess pieces	26.2 chess pieces	32.6 chess pieces	39.7 chess pieces	43.2 chess pieces

The 5-number summary for beginners is:

Min	Q1	Median	Q3	Max
32.5 chess pieces	39.1 chess pieces	48.4 chess pieces	55.7 chess pieces	57.7 chess pieces

The 5-number summary for tournament palyers is:

Min	Q1	Median	Q3	Max
40.1 chess pieces	51.2 chess pieces	64.6 chess pieces	75.9 chess pieces	85.3 chess pieces

c. The range for non-players is:

$$\begin{aligned} Range &= Max - Min \\ &= 43.2 - 22.1 \\ &= 21.2 \text{ chess pieces} \end{aligned}$$

The interquartile range (IQR) for non-players is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 39.7 - 26.2 \\ &= 13.5 \text{ chess pieces} \end{aligned}$$

The range for beginners is:

$$\begin{aligned} Range &= Max - Min \\ &= 57.7 - 32.5 \end{aligned}$$

$$= 25.2 \text{ chess pieces}$$

The interquartile range (IQR) for beginners is:

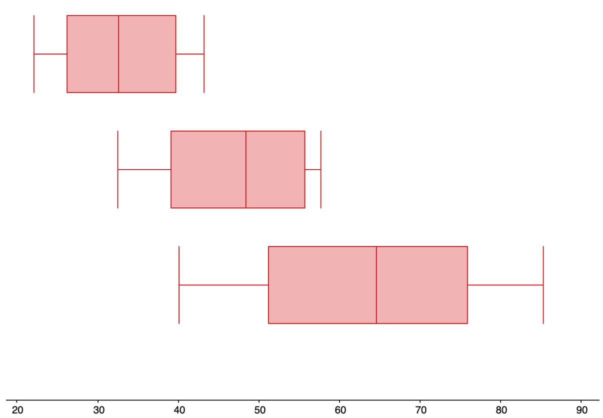
$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 55.7 - 39.1 \\ &= 16.6 \text{ chess pieces} \end{aligned}$$

The range tournament players for is:

$$\begin{aligned} Range &= Max - Min \\ &= 85.3 - 40.1 \\ &= 45.2 \text{ chess pieces} \end{aligned}$$

The interquartile range (IQR) for tournament players is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 75.9 - 51.2 \\ &= 24.7 \text{ chess pieces} \end{aligned}$$

d. 

Tournament players did the best at remembering positions (as shown by all of the numbers of their 5-number summary being larger than the corresponding numbers for the other two groups). However, tournaments players were not completely superior to the other two groups; the best non-players remembered more chess pieces than the worst tournament players. Also tournament players had more variation in how much they.

3.4.10.10.

a. In GeoGebra:

In GeoGebra Classic, enter the data values for false smile, felt smile, miserable smile, and neutral control into columns A, B, C, and D, respectively, of the spreadsheet. Then use the “Multiple Variable Analysis” function. Then use the “Show Statistics” function to display the sample standard deviation for each set of data values.

The sample standard deviation for the false smile group is 1.827. The sample standard deviation for the felt smile group is 1.681. The sample standard deviation for tournament players is 1.454. The sample standard deviation for the neutral control is 1.523.

b. In GeoGebra:

In GeoGebra Classic, enter the data values for false smile, felt smile, miserable smile, and neutral control into columns A, B, C, and D, respectively, of the spreadsheet. Then use the "Multiple Variable Analysis" function. Then use the "Show Statistics" function to display the sample standard deviation for each set of data values.

The 5-number summary for the false smile group is:

Min	Q1	Median	Q3	Max
2.5	3.5	5.5	6.5	9

The 5-number summary for the felt smile group is:

Min	Q1	Median	Q3	Max
2.5	3.5	4.75	6	9

The 5-number summary for the miserable smile group is:

Min	Q1	Median	Q3	Max
2.5	4	4.75	5.5	8

The 5-number summary for the neutral control group is:

Min	Q1	Median	Q3	Max
2	3	4	5	8

c. The range for false smile is:

$$\begin{aligned} Range &= Max - Min \\ &= 9 - 2.5 \\ &= 6.5 \end{aligned}$$

The interquartile range (IQR) for false smile is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 6.5 - 3.5 \\ &= 3 \end{aligned}$$

The range for felt smile is:

$$\begin{aligned} Range &= Max - Min \\ &= 9 - 2.5 \\ &= 6.5 \end{aligned}$$

The interquartile range (IQR) for felt smile is:

$$\begin{aligned} IQR &= Q_3 - Q_1 \\ &= 6 - 3.5 \\ &= 2.5 \end{aligned}$$

The range for miserable smile is:

$$
\begin{aligned}
Range &= Max - Min \\
&= 8 - 2.5 \\
&= 5.5
\end{aligned}
$$

The interquartile range (IQR) for miserable smile is:

$$
\begin{aligned}
IQR &= Q_3 - Q_1 \\
&= 5.5 - 4 \\
&= 1.5
\end{aligned}
$$

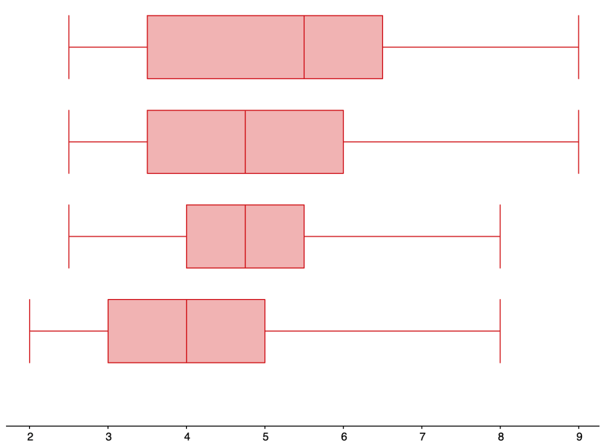
The range for neutral control is:

$$
\begin{aligned}
Range &= Max - Min \\
&= 8 - 2 \\
&= 6
\end{aligned}
$$

The interquartile range (IQR) for neutral control is:

$$
\begin{aligned}
IQR &= Q_3 - Q_1 \\
&= 5 - 3 \\
&= 2
\end{aligned}
$$

d.



The minimum and median for all the smile groups was larger than the minimum and median for the neutral control, indicating that a smile has some effect on leniency. The false smile seems to have been the most effective because the minimum, median, 3rd quartile, and maximum for the false group are greater than or equal to the corresponding values for any of the other groups.

3.4.10.11.

a. There are many possible answers for this question. The data sets $10, 10, 10, 10, 10$ and $9, 9, 10, 11, 11$ have the same mean (10) but different standard deviations (0 and 1, respectively).

b. There are many possible answers for this question. The data sets $2, 2, 2, 2, 2$ and $9, 9, 9, 9, 9$ have the same standard deviation (0) but different means (2 and 9, respectively).

3.4.10.12.

a. There are many possible answers for this question. The data sets $1, 2, 3, 4, 5, 6, 7$ and $11, 12, 13, 14, 15, 16, 17$ have the same IQR (4) but different medians (4 and 14, respectively)

b. There are many possible answers for this question. The data sets $14, 14, 14, 14, 14, 14, 14$ and $1, 1, 1, 14, 26, 26, 26$ have the same median (14) but different IQRs (0 and 25, respectively).

3.4.10.13.

a. The 25th, 50th, and 75th percentiles are, respectively, the 1st quartile, median, and 3rd quartile for the data sets. Reading the boxplot for CPAs, the 25th, 50th, and 75th percentiles for CPAs' salaries are, respectively, $40,000, $75,000, and $90,000. Reading the boxplot for actuaries, the 25th, 50th, and 75th percentiles for actuaries' salaries are, respectively, $75,000, $90,000, and $94,000

b. Deshawn's salary (the median salary for an actuary) is $90,000; Kelsey's salary (the first quartile salary) is also $75,000. So Deshawn makes more than Kelsey, by $15,000.

c. 75% of actuaries make more than the median salary of a CPA ($75,000).

d. 25% of all CPAs earn less than all actuaries.

3.4.10.14.

a. The 25th, 50th, and 75th percentiles for weekly study times for the juniors are 1 hour, 3 hours, and 4 hours, respectively. The 25th, 50th, and 75th percentiles for weekly study times the seniors are 5.5 hours, 6 hours, and 9 hours.

b. Olivia studies more each week than Lucy, by 30 minutes.

c. 50% of juniors study between the minimum and median number of hours for seniors.

d. 100% of seniors study more than the third quartile weekly study time for juniors.

3.4.10.15.

a.

$$\begin{aligned} Z &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{21.4 - 25}{1.15} \\ &= -3.13 \text{ standard deviations} \end{aligned}$$

b. The Z-score for the gas mileage of the car is -3.13 standard deviations.

3.4.10.16.

a.

$$\begin{aligned} Z &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{170 - 274}{63} \\ &= -1.65 \text{ standard deviations} \end{aligned}$$

The marathon time's Z-score is -1.65 standard deviations.

b. Because the marathon finishing time is within 2 standard deviations of the mean finishing time, no, this marathon finishing time is not usually fast.

3.4.10.17.

a. In GeoGebra:

In GeoGebra Classic, enter the data values into the column A of the spreadsheet and use the “One Variable Analysis” function. Then use the “Show Statistics” option to find the mean and standard deviation.

The mean is 46.2 hours per year, and the standard deviation is 6.16 hours per year.

b.

$$\begin{aligned} Z &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{42 - 46.2}{6.16} \\ &= -0.68 \text{ standard deviations} \end{aligned}$$

The Z-score for a city with an average delay time of 42 hours per year is -0.68 standard deviations.

3.4.10.18.

a. In GeoGebra:

In GeoGebra Classic, enter the data values into the column A of the spreadsheet and use the “One Variable Analysis” function. Then use the “Show Statistics” option to find the mean and standard deviation.

The mean is 122.9 job applicants per job posting, and the standard deviation is 10.91 job applicants per job posting.

b.

$$\begin{aligned} Z &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{143 - 122.9}{10.91} \\ &= 1.84 \text{ standard deviations} \end{aligned}$$

The Z-score for a company with 143 job applicants per job posting is 1.84 standard deviations

3.4.10.19.

$$\begin{aligned} Z_{Math} &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{89 - 75}{7} \\ &= 2 \text{ standard deviations} \end{aligned}$$

$$\begin{aligned} Z_{English} &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{65 - 53}{4} \\ &= 3 \text{ standard deviations} \end{aligned}$$

Because the Z-score of my English test is greater than the Z-score of my math test, I did better on the English test than I did on the math test.

3.4.10.20.

$$\begin{aligned} Z_{comedy} &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{102 - 139}{39.7} \\ &= -0.93 \text{ standard deviations} \end{aligned}$$

$$\begin{aligned} Z_{action} &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{129 - 159}{26.2} \\ &= -1.14 \text{ standard deviations} \end{aligned}$$

Because the Z-score for the 129 minute action movie was less than the Z-score for the 102 comedy movie, the action movie was shorter than the comedy movie, when both movies are compared to other movies in their genres.

3.4.10.21.

$$\begin{aligned} Z_{Poe} &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{20.2 - 16.5}{1.85} \\ &= 2 \text{ standard deviations} \end{aligned}$$

$$\begin{aligned} Z_{Gibson} &= \frac{\text{data value} - \text{mean}}{\text{standard deviation}} \\ &= \frac{107 - 81}{13} \\ &= 2 \text{ standard deviations} \end{aligned}$$

Because the Z-scores for the heights of Poe (the Clydesdale horse) and Gibson (the Great Dane) are the same, neither animal is taller than the other when compared to their respective breeds.

3.5 · Chapter 3 Review

· Review Exercises

3.5.1.

a. The population is PCC students and 73,000 is the sample size.

b. The sample is 250 students from the 4 main campuses.

c. Stratified Sample

d. Categorical or Qualitative

e. $435/1000 = 0.435 = 43.5\%$

3.5.2.

a. The population being studied is American adults.

b. The sample is 500 American adults.

c. Categorical or Qualitative

d. The 62% reported in the problem an example of a statistic because it comes from a sample.

e. $62\% - 4\% = 58\%$ and $62\% + 4\% = 66\%$. The confidence interval is $(58\%, 66\%)$ and it is in relation to the parameter.

f. We are confident that the true proportion of all adult Americans who favor a law to ban the sale of assault weapons and semi automatic rifles is between 58% and 66%.

3.5.3.

a. The population being studied is PCC Students.

b. Categorical or Qualitative

c. 23% is a statistic because it is from a sample.

d. The margin of error is 4%.

e. $23\% - 4\% = 19\%$ and $23\% + 4\% = 27\%$. The confidence interval is $(19\%, 27\%)$.

f. We are confident that the true proportion of all PCC students who prefer to study at the library is between 19% and 27%

3.5.4.

a. Systematic

b. Simple Random Sample

c. Convenience

d. Systematic

3.5.5.

a. Voluntary Response Bias

b. Sampling Bias

c. Sampling Bias

d. Response bias

e. Perceived lack of anonymity

3.5.6.

a. Observational

b. Experiment

3.5.7.

a. The treatment group is the group receiving the test medicine for migraines.

b. The control group is the group receiving the inert pill.

c. This is a blind study.

d. This is a placebo-controlled experiment.

3.5.8.

a. The treatment group is the group assigned to use the Rewire app.

b. The control group is the group that receives the usual services from the Department of Youth Services.

c. This study is neither blind nor double-blind because the researchers and the youth know whether they are using the Rewire app of receiving the usual services from the Department of Youth Services.

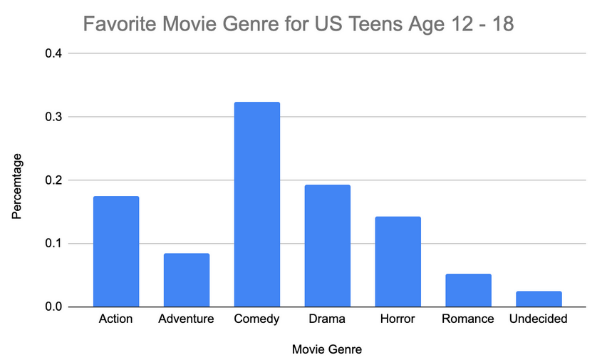
d. This is a controlled experiment.

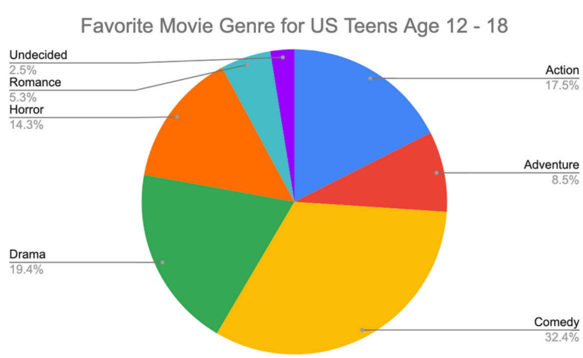
3.5.9.

a. The implied population is United States teenagers.

b. There were 2007 teens surveyed.

c. Qualitative or Categorical

d. 

e. 

f. Answers will vary

g. About 14.3% of teens chose horror as their favorite movie genre.

3.5.10.

a. 426 people use TriMet for personal business.

b. 479 people people use TriMet to get to the airport.

3.5.11.

a. The mean is \$4.78. The median is \$4.75.

b. The mean and median is about the same value therefore the data is symmetric.

c. The standard deviation is \$1.34.

d. $z_{\$3.25} = \frac{\$3.25 - \$4.78}{\$1.34} = \frac{-\$1.53}{\$1.34} = -1.14$

$z_{\$8.95} = \frac{\$8.95 - \$4.78}{\$1.34} = \frac{\$4.17}{\$1.34} = 3.11$

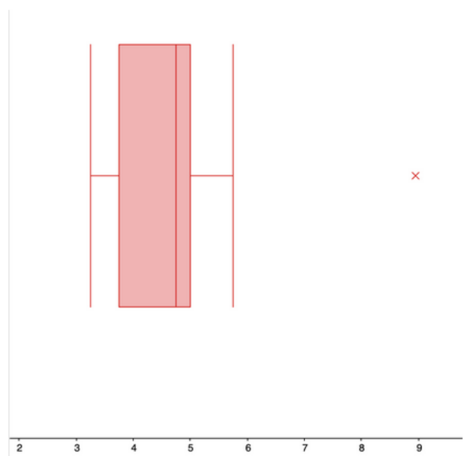
e.

Min	3.25
Q1	3.75
Median	4.75
Q3	5
Max	8.95

f. Range $= \$8.95 - \$3.25 = \$5.70$

IQR $= \$5 - \$3.75 = \$1.25$

g.



3.5.12.

a. The mean is about 78.9.

b. The median is 78.

c. The mean is greater than the median therefore the data is left skewed.

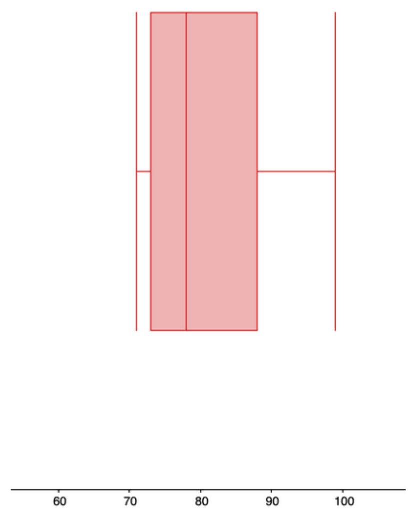
d. The standard deviation is about 14.5.

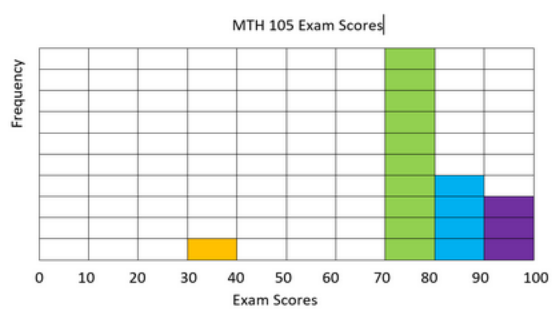
e. The majority of values are between 64.4 and 93.4.

f. $z = \frac{(99-78.9)}{14.5} = 1.39$. This 99 is only 1.39 standard deviations above the mean. This is not unusual. Generally, a value should be more than three deviations above or below the mean to be considered unusual.

g.

Min	32
Q1	73
Median	78
Q3	88
Max	99

h. 

i. 

3.5.13.

a. The mean is about \$26,200 and the standard deviation is about \$10,000.

b. The majority of cars sell for between \$16,200 and \$36,200.

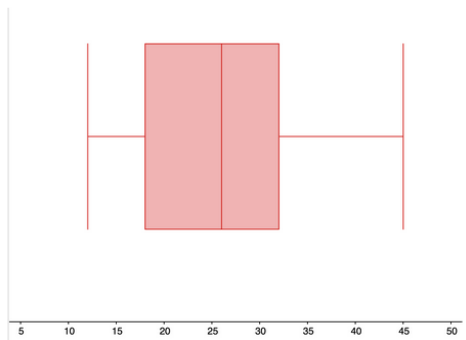
c.

Min	12
Q1	18
Median	26
Q3	32
Max	45

d. Range $= 45 - 12 = 33$

IQR $= 32 - 18 = 14$

e.



3.5.14.

a. The 25th, 50th, and 75th percentile for Team A goals are 2, 4, and 5, respectively.

The 25th, 50th, and 75th percentile for Team B goals are 6, 8, and 9, respectively.

b. The Median for Team A is 4.

The Median for Team B is 8.

c. Fifty percent of the goals for Team B exceed the maximum number for goals for Team A.

d. The data for Team A is more symmetric than for Team B.

e. The data for Team B is skewed left.

3.5.15.

a. $z = \dfrac{(30.4 - 35)}{1.35} = \dfrac{-4.6}{1.35} \approx -3.41$

b. 30.4 mpg is 3.41 standard deviations below the mean so the car is getting unusually low gas mileage.

3.5.16.

a. The mean is approximately 50.6 minutes per week. The standard deviation approximately 12.2 minutes per week.

b. $z = \dfrac{(42 - 50.6)}{12.2} \approx -0.70$

c. 42 minutes per week is 0.70 standard deviations below the mean, so it is not unusual.

3.6 · The Normal Distribution

3.6.8 · Exercises

3.6.8.1.

The shape of a Normal distribution is unimodal, symmetric and bell-shaped.

3.6.8.2.

50% of the observations will be below the mean.

3.6.8.3.

The median tree diameter is approximately 35 inches.

3.6.8.4.

Distribution B has a wider spread because it has a larger standard deviation.

3.6.8.5.

Approximately 68% of the values fall within one standard deviation of the mean.

3.6.8.6.

Approximately 95% of the values fall within two standard deviations of the mean.

3.6.8.7.

Approximately 99.7% of the values fall within three standard deviations of the mean.

3.6.8.8.

Approximately 0.3% of the values fall outside of three standard deviations of the mean.

3.6.8.9.

Approximately 27% of the values fall between the first and second standard deviations from the mean.

3.6.8.10.

Approximately 31.7% of the values fall between the first and third standard deviations from the mean.

3.6.8.11.

Approximately 4.7% of the values fall between the second and third standard deviations from the mean.

3.6.8.12.

Approximately 32% of the values fall outside the first standard deviation from the mean.

3.6.8.13.

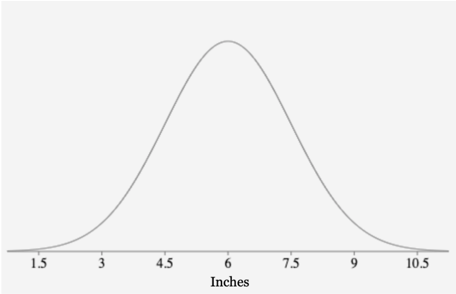
Approximately 34% of the values fall between the mean and one standard deviation below the mean.

3.6.8.14.

Approximately 49.85% of the values fall between the mean and three standard deviations above the mean.

3.6.8.15.

a. Here is the graph:



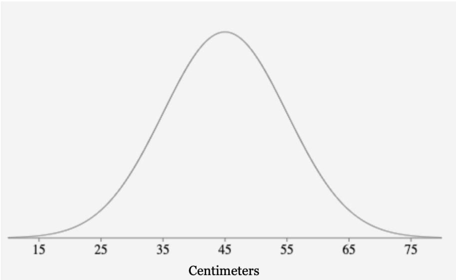
b. Data values from 4.5 to 6.5 inches fall within one standard deviation of the mean.

c. The percentage of data that fall between 3 and 10.5 inches is $13.5\% + 34\% + 34\% + 13.5\% + 2.35\% = 97.35\%$.

d. The percentage of data that fall below 1.5 inches is 0.15%.

3.6.8.16.

a. Here is the graph:



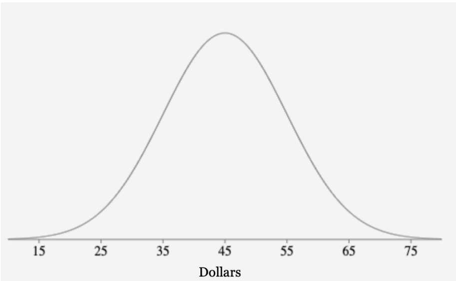
b. The range of data values that fall within two standard deviations of the mean is 25 to 65 cm.

c. The percentage of the data that fall between 15 and 55 cm is $2.35\% + 13.5\% + 34\% + 34\% = 83.85\%$.

d. The percentage of the data that fall above 55 cm is $13.5\% + 2.35\% + 0.15\% = 16\%$.

3.6.8.17.

a. Here is the graph:



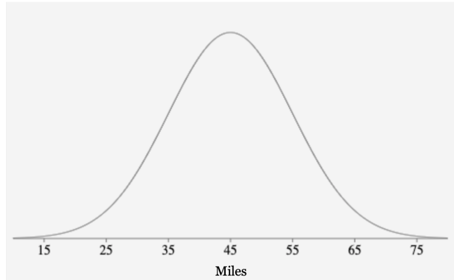
b. The range of data values that fall within three standard deviations of the mean is \$4 to \$16.

c. The percentage of data that lie between \$6 and \$14 is 95%.

d. The percentage of data that lie above \$14 is $2.35\% + 0.15\% = 2.5\%$.

3.6.8.18.

a. Here is the graph:



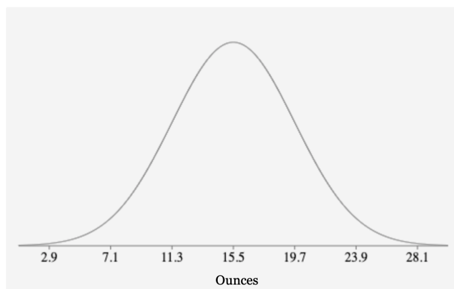
b. The range of data values that falls within one standard deviation of the mean is 12 to 18 miles.

c. The percentage of the data that fall between 9 and 18 miles is $13.5\% + 34\% + 34\% = 81.5\%$.

d. The percentage of the data that fall above 18 or below 9 miles is $100\% - 81.5\% = 18.5\%$.

3.6.8.19.

a. Here is the graph:



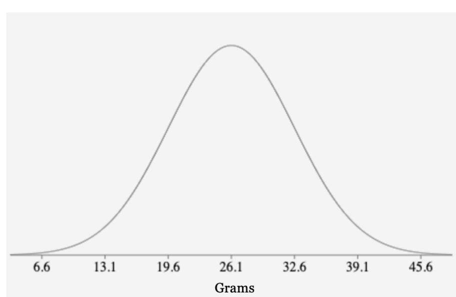
b. The percentage of the data values that lie above 18.6 ounces is 0.2302 or 23.028%.

c. The percentage of the data values that lie between 9 and 20.2 ounces is 0.8076 or 80.76%.

d. The percentage of the data values that lie below 13.7 ounces is 0.3341 or 33.41%.

3.6.8.20.

a. Here is the graph:



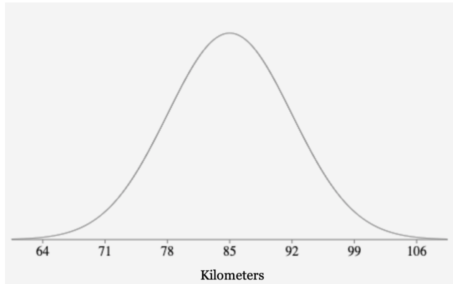
b. The percentage of the data values that fall above 32.6 grams is 0.1587 or 15.87%.

c. The percentage of the data values that is below 15 grams or greater than 36.7 grams is 0.0953 or 9.53%.

d. The percentage of the data values that is less than or equal to 20.8 grams is 0.2074 or 20.74%.

3.6.8.21.

a. Here is the graph:



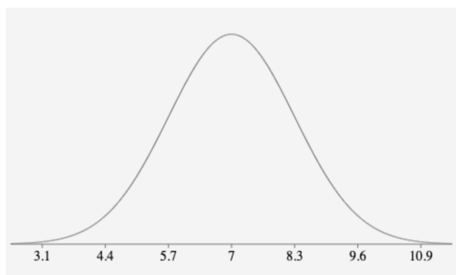
b. $P(X \leq 82) = 0.3341$.

c. $P(76 \leq X \leq 90) = 0.6632$.

d. $P(X \geq 100) = 0.0161$.

3.6.8.22.

a. Here is the graph:



b. $P(X \geq 8) = 0.2209$.

c. $P(X \leq 4.1)$ or $P(X \geq 7.8) = 0.282$.

d. $P(X \leq 7.3) = 0.5913$.

3.6.8.23.

A z-score measures how many standard deviations from the mean a data value is.

3.6.8.24.

For a Standard Normal distribution, the mean is always 0 and the standard deviation is always 1.

3.6.8.25.

$Z = \frac{6.2-6}{1.5} \approx= 0.13$ standard deviations.

3.6.8.26.

$Z = \frac{32-45}{10} = -1.3$ standard deviations.

3.6.8.27.

$Z = \frac{5-10}{2} = -2.5$ standard deviations.

3.6.8.28.

$Z = \frac{19-15}{3} \approx= 1.13$ standard deviations.

3.6.8.29.

The confidence interval is $(\$2.20, \$2.50)$. This means that we are 95% confident that the true average amount of change for all those who carry a purse is between \$2.20 and \$2.50.

3.6.8.30.

The confidence interval is $(12, 16)$. This means that we are 95% confident that the true average quiz score is between 12 and 16 points.

3.6.8.31.

The margin of error is 0.2156 cm. The confidence interval is approximately $(82.28, 84.72)$. This means that we are 95% confident that the true population parameter is between approximately 82.28 and 84.72 cm.

3.6.8.32.

The margin of error is \$1.47. The confidence interval is $(\$33.93, \$36.87)$. This means that we are 95% confident that the true population parameter is between \$33.93 and 36.87.

4 · Probability

4.1 · Contingency Tables

4.1.10 · Exercises

4.1.10.1.

	Food Insecure	Not Food Insecure	Total
Housing Insecure	***380***	60	***440***
Not Housing Insecure	***300***	460	760
Total	680	***520***	***1200***

4.1.10.2.

	Bookstore	No Bookstore	Total
Cafeteria	***290***	***85***	375
No Cafeteria	***340***	135	***475***
Total	630	***220***	850

4.1.10.3.

	Breakfast	No Breakfast	Total
Floss	12	49	61
No Floss	3	8	11
Total	15	57	72

4.1.10.4.

	Chromebook	No Chromebook	Total
Apple	65	120	185
No Apple	85	45	130
Total	150	165	315

4.1.10.5.

	A	Not A	Total
B	10	20	30
Not B	20	25	45
Total	30	45	75

4.1.10.6.

	A	Not A	Total
B	30	50	80
Not B	30	10	40
Total	60	60	120

4.1.10.7.

a.

$$\begin{aligned} \text{P(In morning class)} &= \frac{39}{65} \\ &= 0.60 \text{ or } 60\% \end{aligned}$$

b.

$$\begin{aligned} \text{P(Earned a C)} &= \frac{25}{65} \\ &\approx 0.385 \text{ or } 38.5\% \end{aligned}$$

c.

$$\begin{aligned} \text{P(Earned an A and in afternoon class)} &= \frac{10}{65} \\ &\approx 0.154 \text{ or } 15.4\% \end{aligned}$$

d.

$$\begin{aligned} \text{P(Earned an A given in morning class)} &= \frac{8}{39} \\ &\approx 0.205 \text{ or } 20.5\% \end{aligned}$$

e.

$$\begin{aligned} \text{P(In morning class or earned B)} &= \frac{43}{65} \\ &\approx 0.662 \text{ or } 66.2\% \end{aligned}$$

4.1.10.8.

a.

$$\begin{aligned} \text{P(in morning class)} &= \frac{32}{60} \\ &\approx 0.533 \text{ or } 53.3\% \end{aligned}$$

b.

$$\begin{aligned} \text{P(Freshman)} &= \frac{17}{60} \\ &\approx 0.283 \text{ or } 28.3\% \end{aligned}$$

c.

$$\begin{aligned} \text{P(Senior and in afternoon class)} &= \frac{2}{60} \\ &\approx 0.033 \text{ or } 3.3\% \end{aligned}$$

d.

$$\begin{aligned} \text{P(Sophomore given in morning class)} &= \frac{5}{32} \\ &\approx 0.156 \text{ or } 15.6\% \end{aligned}$$

e.

$$\begin{aligned} \text{P(in morning class or Junior)} &= \frac{40}{60} \\ &\approx 0.667 \text{ or } 66.7\% \end{aligned}$$

4.1.10.9.

a.

$$\begin{aligned} \text{P(no credit cards)} &= \frac{27}{81} \\ &\approx 0.333 \text{ or } 33.3\% \end{aligned}$$

b.

$$\begin{aligned} \text{P(one credit card)} &= \frac{15}{81} \\ &\approx 0.185 \text{ or } 18.5\% \end{aligned}$$

c.

$$\begin{aligned} \text{P(no credit cards and over age 35)} &= \frac{18}{81} \\ &\approx 0.222 \text{ or } 22.2\% \end{aligned}$$

d.

$$\begin{aligned} \text{P(between ages of 18 and 35,or have zero credit cards)} &= \frac{51}{81} \\ &\approx 0.630 \text{ or } 63.0\% \end{aligned}$$

e.

$$\begin{aligned} \text{P(no credit cards given between ages of 18 and 35)} &= \frac{9}{33} \\ &\approx 0.273 \text{ or } 27.3\% \end{aligned}$$

f.

$$\begin{aligned} \text{P(no credit cards given over age 35)} &= \frac{18}{48} \\ &= 0.375 \text{ or } 37.5\% \end{aligned}$$

g. Yes, it appears that having no credit cards depends on age. The probability of having no credit cards for people over age 35 is significantly greater than the probability of having no credits for people between the ages of 18 and 35.

4.1.10.10.

a.

$$\begin{aligned} P(\text{inoculated}) &= \frac{244}{6224} \\ &\approx 0.039 \text{ or } 3.9\% \end{aligned}$$

b.

$$\begin{aligned} P(\text{lived}) &= \frac{5374}{6224} \\ &\approx 0.863 \text{ or } 86.3\% \end{aligned}$$

c.

$$\begin{aligned} P(\text{ied given or inoculated}) &= \frac{1088}{6224} \\ &\approx 0.175 \text{ or } 17.5\% \end{aligned}$$

d.

$$\begin{aligned} P(\text{ied given inoculated}) &= \frac{6}{244} \\ &\approx 0.025 \text{ or } 2.5\% \end{aligned}$$

e.

$$\begin{aligned} P(\text{died given not inoculated}) &= \frac{844}{5980} \\ &\approx 0.141 \text{ or } 14.1\% \end{aligned}$$

f. Yes, it appears that survival was dependent if a person was inoculated. The percentage of deaths among the not inoculated group was nearly six times greater than the percentage of deaths among the inoculated group.

4.1.10.11.

a.

$$\begin{aligned} P(\text{not survive}) &= \frac{1490}{2201} \\ &\approx 0.677 \text{ or } 67.7\% \end{aligned}$$

b.

$$\begin{aligned} P(\text{crew}) &= \frac{885}{2201} \\ &\approx 0.402 \text{ or } 40.2\% \end{aligned}$$

c.

$$\begin{aligned} P(\text{first class and not survive}) &= \frac{122}{2201} \\ &\approx 0.055 \text{ or } 5.5\% \end{aligned}$$

d.

$$\begin{aligned}P(\text{not survive or crew}) &= \frac{1702}{2201}\\ &\approx 0.773 \text{ or } 77.3\%\end{aligned}$$

e.

$$\begin{aligned}P(\text{survived given first class}) &= \frac{203}{325}\\ &\approx 0.625 \text{ or } 62.5\%\end{aligned}$$

f.

$$\begin{aligned}P(\text{survived given second class}) &= \frac{118}{285}\\ &\approx 0.414 \text{ or } 41.4\%\end{aligned}$$

g.

$$\begin{aligned}P(\text{survived given third class}) &= \frac{178}{706}\\ &\approx 0.252 \text{ or } 25.2\%\end{aligned}$$

h. Yes, it does appear that survival depended on the passenger's class. The probability of survival for first class passengers is significantly greater than the probability of survival for second class passengers and is more than double the probability of survival for third class passengers.

4.1.10.12.

a.

$$\begin{aligned}P(\text{foreign and government}) &= \frac{1\%}{100\%}\\ &= 0.01 \text{ or } 1\%\end{aligned}$$

b.

$$\begin{aligned}P(\text{U.S. and corporation}) &= \frac{45\%}{100\%}\\ &= 0.45 textor 45\%\end{aligned}$$

c.

$$\begin{aligned}P(\text{foreign or government}) &= \frac{47\%}{100\%}\\ &= 0.47 \text{ or } 47\%\end{aligned}$$

d.

$$\begin{aligned}P(\text{U.S. given individual}) &= \frac{8\%}{11\%}\\ &\approx 0.727 \text{ or } 72.7\%\end{aligned}$$

e.

$$\begin{aligned}P(\text{foreign given government}) &= \frac{1\%}{3\%}\\ &\approx 0.333 \text{ or } 33.3\%\end{aligned}$$

4.1.10.13.

a.

	Game/Software	No Game/Software	Total
Computer	10%	5%	15%
No Computer	15%	70%	85%
Total	25%	75%	100%

b.

$$\begin{aligned}\text{P(no computer and no game/software)} &= \frac{70\%}{100\%}\\ &= 0.7 \text{ or } 70\%\end{aligned}$$

c.

$$\begin{aligned}\text{P(computer or game/software)} &= \frac{30\%}{100\%}\\ &= 0.3 \text{ or } 30\%\end{aligned}$$

d.

$$\begin{aligned}\text{P(game/software given computer)} &= \frac{10\%}{15\%}\\ &\approx 0.667 \text{ or } 66.7\%\end{aligned}$$

e.

$$\begin{aligned}\text{P(game/software given no computer)} &= \frac{15\%}{85\%}\\ &\approx 0.176 \text{ or } 17.6\%\end{aligned}$$

f. Purchasing a game/software and purchasing a computer appear to be depended. The probability of purchasing a game/software for computer buyers was almost 50% greater than the probability of purchasing a game/software among customers who did not purchase a computer.

4.1.10.14.

a.

	Injury	No Injury	Total
Stretched	52	270	322
Not Stretched	21	57	78
Total	73	327	400

b.

$$\begin{aligned}\text{P(injury)} &= \frac{73}{400}\\ &\approx 0.183 \text{ or } 18.3\%\end{aligned}$$

c.

$$\begin{aligned}\text{P(injury and did not stretch)} &= \frac{21}{400}\\ &\approx 0.053 \text{ or } 5.3\%\end{aligned}$$

d.

$$\text{P(stretched or no injury)} = \frac{379}{400}$$

$$\approx 0.948 \text{ or } 94.8\%$$

e.

$$\begin{aligned} \text{P(injury give stretched)} &= \frac{52}{322} \\ &\approx 0.161 \text{ or } 16.1\% \end{aligned}$$

f.

$$\begin{aligned} \text{P(injury give did not stretch)} &= \frac{21}{78} \\ &\approx 0.269 \text{ or } 26.9\% \end{aligned}$$

g. It apppears that sustaining an injury is dependent on whether the member stretches before exercising. The probability of sustaining an injury among members who did not stretch before exercising was significantly higher than the probability of sustaining an injury among members who stretched before exercising.

4.1.10.15.

a.

	Hardcover	Paperback	Total
Fiction	13	59	72
Nonfiction	15	8	23
Total	28	67	95

b.

$$\begin{aligned} \text{P(non-fiction and paperback)} &= \frac{8}{95} \\ &\approx 0.084 \text{ or } 8.4\% \end{aligned}$$

c.

$$\begin{aligned} \text{P(fiction given hardcover)} &= \frac{13}{28} \\ &\approx 0.464 \text{ or } 46.4\% \end{aligned}$$

4.1.10.16.

a.

	Contingency Table	No Contingency Table	Total
Pass	23	4	27
No Pass	2	3	5
Total	25	7	32

b.

$$\begin{aligned} \text{P(pass and no contingency table)} &= \frac{23}{32} \\ &\approx 0.719 \text{ or } 71.9\% \end{aligned}$$

c.

$$\begin{aligned} \text{P(pass given no contingency table)} &= \frac{4}{7} \\ &\approx 0.571 \text{ or } 57.1\% \end{aligned}$$

4.2 · Theoretical Probability

4.2.15 · Exercises

4.2.15.1.

a. 2/13

b. 6/13

c. 0

d. 25/169

e. 1/13

4.2.15.2.

a. 1/2

b. 1/3

4.2.15.3.

a. 11/12

b. 1/6

4.2.15.4.

a. 1/36

b. 1/12

4.2.15.5.

6241/10000

4.2.15.6.

a. 1/60

b. 1/4

4.2.15.7.

a. 1/8

b. 1/8

4.2.15.8.

a. 7/12

b. 2/3

4.2.15.9.

a. 1/4

b. 1/6

c. $1/9$

4.2.15.10.

a. $1/7776$

b. $1 - \text{P(No 3's)} = 4651/7776$

4.2.15.11.

a. $4/52 = 1/13$

b. $13/52 = 1/4$

c. $26/52 = 1/2$

d. $16/52 = 4/13$

4.2.15.12.

$1 - \text{P(no dark chocolates)} = 84/90 = 14/15 \approx 0.933$ or 93.3%

4.2.15.13.

$60/1320 = 1/22 \approx 0.045$ or 4.5%

4.2.15.14.

$125/1728 \approx 0.072$ or 7.2%

4.2.15.15.

$1 - \text{P(no blue marbles)} = 1110/1320 = 37/44 \approx 0.841$ or 84.1%

4.3 · Expected Value

4.3.2 · Exercises

4.3.2.1.

a.

Die roll	Gold	Silver	Black
Outcome	\$3	\$2	-\$1>
Probability	3/37	6/37	28/37

b. $3(3/37) + 2(6/37) - 1(28/37) = -0.19$

The expected value is approximately -\$0.19. That is, you would lose about \$0.19 on average each time you pick a marble.

4.3.2.2.

a.

Die roll	Red	Blue	Green
Outcome	\$3	\$2	-\$1
Probability	5/28	8/28	15/28

b. $3(5/28) + 2(8/28) - 1(15/28) = 0.57$

The expected value is approximately \$0.57. That is, you would win about \$0.57 on average each time you pick a marble.

4.3.2.3.

a.

Die roll outcome	1, 2, 3, or 4	5	6
Outcome	\$5	\$0	-\$2
Probability	1/6	1/6	4/6

b. $5(1/6) + 0(1/6) - 2(4/6) = -0.50$

The expected value is about -\$0.50 which means you would lose 50 cents on average each time you roll the die.

c. No, you should not play this game (unless you want to give your friend your money.

4.3.2.4.

a.

Die roll outcome	1	2	3, 4, 5, or 6
Probability	4/6	1/6	1/6

b. The expected value is -\$0.17 which means you would lose 17 cents on average each time you roll the die.

c. No, you should not play this game (unless you want to give your friend your money.

4.3.2.5.

The company's expected profit is \$45.55 per warranty sold.

4.3.2.6.

The company's expected profit is \$58 per warranty sold.

4.3.2.7.

The company's expected value on each policy is \$22 which means they will make \$22, on average, per policy sold.

4.3.2.8.

The company's expected value on each policy is \$2,986 which means they will make \$2,986, on average, per policy sold.

4.3.2.9.

The expected value for this raffle is -\$3.

4.3.2.10.

The expected value for this raffle is -\$7.

4.3.2.11.

The expected value for this game is approximately -\$6.

4.3.2.12.

The expected value for this game is approximately -\$7.19.

4.3.2.13.

Answers will vary since you are making up your own problem.

4.4 · Chapter 4 Review

· Review Exercises

4.4.1.

a. P(in the afternoon class) $\approx 28/60 = 0.4667$

b. P(earned an A) $25/60 \approx 0.4167$

c. P(earned a B and was in the afternoon class) $= 13/60 \approx 0.2167$

d. P(earned a C given the student was in the morning class) $= 7/32 \approx 0.2188$

e. P(is in the morning class given that the student earned a B) $= 11/24 \approx 0.4583$

4.4.2.

a. P(in the math class) $= 27/65 \approx 0.4154$

b. P(earned a B) $= 26/65 = 0.40$

c. P(earned an A and was in the math class) $= 10/65 \approx 0.1538$

d. P(earned a B given the student was in the science class) $= 18/38 \approx 0.4737$

e. P(is in the math class given that the student earned a C) $= 8/26 \approx 0.3077$

4.4.3.

Type	Channel 2	Channel 6	Channel 8	Channel 12	Total
Drama	5	2	4	***4***	15
Sitcom	6	***9***	7	3	25
Game Show	***4***	4	3	4	15
News	3	2	***2***	3	10
Total	***18***	***17***	***16***	***14***	65

a. P(Sitcom or Game Show) $= 40/65 \approx 0.6154$

b. P(Drama and Channel 8) $= 4/65 \approx 0.0615$

c. P(Channel 8 or Channel 2) $= 34/65 \approx 0.5231$

d. P(Drama given that it is on Channel 6) $= 2/17 \approx 0.1176$

e. P(Channel 12 given that it's a sitcom) $= 3/25 = 0.12$

f. P(Game show given that it is on Channel 2) $= 4/18 \approx 0.2222$

4.4.4.

Type	Channel 3	Channel 5	Channel 7	Channel 13	Total
Reality	6	***7***	8	6	***27***
Crime	2	1	***4***	2	***9***
Cooking	5	5	9	11	***30***
Community	***2***	8	4	***3***	***17***
Total	15	21	25	***22***	83

a. P(Reality or Crime Show) $= 36/83 \approx 0.4337$

b. P(Cooking and Channel 3) $= 5/83 \approx 0.0602$

c. P(Channel 3 or Channel 13) $= 37/83 \approx 0.4458$

d. P(Community Program given that it is on Channel 5) $= 8/21 \approx 0.3810$

e. P(Channel 3 given it is a crime show) $= 2/9 \approx 0.2222$

f. P(Cooking show given it is on channel 7) $= 9/25 = 0.36$

4.4.5.

a. P(red) = 12/25 = 0.48
b. P(not white) = 17/25 = 0.68
c. P(Yellow or Red) = 17/25 = 0.68
d. P(Blue) = 0/25 = 0
e. P(Two reds) = 144/625 = 0.2304
f. P(Red then yellow) = 60/600 = 0.10

4.4.6.

a. P(Black) = 18/31 $\approx$ 0.5806
b. P(Not purple) = 27/31 $\approx$ 0.8710
c. P(Blue or purple) = 13/31 $\approx$ 0.4194
d. P(Yellow) = 0/31 = 0
e. P(Two black) = 324/961 $\approx$ 0.3371
f. P(Blue then black) = 162/930 $\approx$ 0.1742

4.4.7.

a. P(Four heads) = 0.0625
b. P(No heads) = 0.0625

4.4.8.

a. P(All tails) $\approx$ 0.0078
b. P(All heads) $\approx$ 0.0078

4.4.9.

a. P(Both too little) $\approx$ 0.4624
b. P(Neither too little) $\approx$ 0.1024

4.4.10.

a. P(All support) $\approx$ 0.8574
b. P(None support) $\approx$ 0.0001

4.4.11.

a.

Color	Black	Orange	Yellow
x	\$3	\$2	-\$1
P(x)	2/26	4/26	20/26

b. The expected value is -\$0.23.
c. No, you should not play this game because it has a negative expected value.

4.4.12.

a.

Roll	6	5	4	3	2	1
x	\$10	\$0	-\$1	-\$1	-\$1	-\$1
P(x)	1/6	1/6	1/6	1/6	1/6	1/6

b. The expected value is \$1.00.

c. Yes, you should expect to win money because this game has a positive expected value.

4.4.13.

Outcome	Product Failed	Didn't Fail
x	-\$450	\$0
P(x)	0.015	0.985

The expected loss per warranty is \$48.25. (Note that if you include the \$55.00 a person pays for a warranty, the expected profit will be \$6.75 per warranty)

4.4.14.

Outcome	Winning Ticket	No Winning Ticket
x	\$8000	\$0
P(x)	1/2000	1999/2000

The expected loss per ticket is \$6.00. So the expected value is -\$6.00. (Note that if you inlucde the \$10 it costs to but the ticket, the expected value is a profit of \$4.00)

5 · Democracy

5.1 · Apportionment

5.1.11 · Exercises

5.1.11.1.

a. Math: 6 tutors, English: 5 tutors, Chemistry: 3 tutors, Biology: 1 tutor

b. Math: 7 tutors, English: 5 tutors, Chemistry: 2 tutors, Biology: 1 tutor, Modified divisor 47

c. Math: 6 tutors, English: 5 tutors, Chemistry: 3 tutors, Biology: 1 tutor, Modified divisor 52

d. Math: 6 tutors, English: 5 tutors, Chemistry: 3 tutors, Biology: 1 tutor, Divisor 53

5.1.11.2.

a. Math: 8 tutors, English: 7 tutors, Chemistry: 3 tutors, Biology: 2 tutors

b. Math: 9 tutors, English: 7 tutors, Chemistry: 3 tutors, Biology: 1 tutor, Modified divisor 36

c. Math: 8 tutors, English: 7 tutors, Chemistry: 3 tutors, Biology: 2 tutors, Divisor 39.75

d. Math: 8 tutors, English: 7 tutors, Chemistry: 3 tutors, Biology: 2 tutors, Divisor 39.75

5.1.11.3.

a. Morning: 1 salesperson, Midday: 5 salespeople, Afternoon: 6 salespeople, Evening: 8 salespeople

b. Morning: 1 salesperson, Midday: 4 salespeople, Afternoon: 7 salespeople, Evening: 8 salespeople, Modified divisor 62

c. Morning: 1 salesperson, Midday: 5 salespeople, Afternoon: 6 salespeople, Evening: 8 salespeople, Divisor 67.5

d. Morning: 1 salesperson, Midday: 5 salespeople, Afternoon: 6 salespeople, Evening: 8 salespeople, Divisor 67.5

5.1.11.4.

a. Morning: 2 salespeople, Midday: 6 salespeople, Afternoon: 8 salespeople, Evening: 9 salespeople

b. Morning: 1 salesperson, Midday: 6 salespeople, Afternoon: 8 salespeople, Evening: 10 salespeople, Modified divisor 50

c. Morning: 2 salespeople, Midday: 6 salespeople, Afternoon: 8 salespeople, Evening: 9 salespeople, Modified divisor 55

d. Morning: 2 salespeople, Midday: 6 salespeople, Afternoon: 8 salespeople, Evening: 9 salespeople, Modified divisor 55

5.1.11.5.

a. Aisha: 18 coins, Basir: 14 coins, Carlos: 4 coins

b. Aisha: 19 coins, Basir: 14 coins, Carlos: 3 coins, Modified divisor 400

c. Aisha: 19 coins, Basir: 14 coins, Carlos: 3 coins, Modified divisor 410

d. Aisha: 19 coins, Basir: 14 coins, Carlos: 3 coins, Modified divisor 410

5.1.11.6.

a. Aisha: 19 coins, Basir: 15 coins, Carlos: 3 coins

b. Aisha: 19 coins, Basir: 15 coins, Carlos: 3 coins, Modified divisor 390

c. Aisha: 19 coins, Basir: 15 coins, Carlos: 3 coins, Divisor 402.7

d. Aisha: 19 coins, Basir: 15 coins, Carlos: 3 coins, Modified divisor 405

5.1.11.7.

a. A: 5 seats, B: 79 seats, C: 20 seats, D: 12 seats

b. A: 4 seats, B: 80 seats, C: 20 seats, D: 12 seats, Modified divisor 6950

c. A: 5 seats, B: 78 seats, C: 20 seats, D: 13 seats, Modified divisor 7128

d. A: 5 seats, B: 78 seats, C: 20 seats, D: 13 seats, Modified divisor 6950

5.1.11.8.

a. A: 5 seats, B: 84 seats, C: 21 seats, D: 14 seats

b. A: 5 seats, B: 85 seats, C: 21 seats, D: 13 seats, Modified divisor 6550

c. A: 5 seats, B: 85 seats, C: 21 seats, D: 13 seats, Modified divisor 6620

d. A: 5 seats, B: 85 seats, C: 21 seats, D: 13 seats, Modified divisor 6610

5.1.11.9.

a. A: 40 seats, B: 24 seats, C: 15 seats, D: 30 seats, E: 10 seats

b. A: 41 seats, B: 24 seats, C: 14 seats, D: 30 seats, E: 10 seats, Modified divisor 19,700

c. A: 40 seats, B: 24 seats, C: 15 seats, D: 30 seats, E: 10 seats, Modified divisor 20,100

d. A: 40 seats, B: 24 seats, C: 15 seats, D: 30 seats, E: 10 seats, Modified divisor 20,125

5.1.11.10.

a. A: 43 seats, B: 25 seats, C: 16 seats, D: 31 seats, E: 11 seats

b. A: 43 seats, B: 25 seats, C: 15 seats, D: 32 seats, E: 11 seats, Modified divisor 18,500

c. A: 43 seats, B: 25 seats, C: 15 seats, D: 32 seats, E: 11 seats, Modified divisor 18,850

d. A: 43 seats, B: 25 seats, C: 15 seats, D: 32 seats, E: 11 seats, Modified divisor 18,850

5.1.11.11.

a. A: 23 seats, B: 16 seats, C: 77 seats, D: 30 seats, E: 21 seats, F: 33 seats

b. A: 22 seats, B: 16 seats, C: 78 seats, D: 30 seats, E: 21 seats, F: 33 seats, Modified divisor 148.5

c. A: 23 seats, B: 16 seats, C: 77 seats, D: 30 seats, E: 21 seats, F: 33 seats, Divisor 150

d. A: 23 seats, B: 16 seats, C: 77 seats, D: 30 seats, E: 21 seats, F: 33 seats, Divisor 150

5.1.11.12.

a. A: 20 seats, B: 15 seats, C: 69 seats, D: 27 seats, E: 19 seats, F: 30 seats

b. A: 20 seats, B: 14 seats, C: 70 seats, D: 27 seats, E: 19 seats, F: 30 seats, Modified divisor 164

c. A: 20 seats, B: 15 seats, C: 69 seats, D: 27 seats, E: 19 seats, F: 30 seats, Modified divisor 166.9

d. A: 20 seats, B: 15 seats, C: 69 seats, D: 27 seats, E: 19 seats, F: 30 seats, Modified divisor 167

5.1.11.13.

a. . A: 19 seats, B: 19 seats, C: 22 seats, D: 22 seats, E: 81 seats, F: 87 seats

b. A: 28 seats, B: 19 seats, C: 22 seats, D: 22 seats, E: 82 seats, F: 87 seats, Modified divisor 4347

c. A: 19 seats, B: 19 seats, C: 22 seats, D: 22 seats, E: 81 seats, F: 87 seats, Divisor 4400.4

d. A: 19 seats, B: 19 seats, C: 22 seats, D: 22 seats, E: 81 seats, F: 87 seats, Divisor 4400.4

5.1.11.14.

a. A: 18 seats, B: 19 seats, C: 21 seats, D: 21 seats, E: 78 seats, F: 83 seats

b. A: 18 seats, B: 18 seats, C: 21 seats, D: 21 seats, E: 78 seats, F: 84 seats, Modified divisor 4550

c. A: 18 seats, B: 18 seats, C: 21 seats, D: 21 seats, E: 78 seats, F: 84 seats, Modified divisor 4580

d. A: 18 seats, B: 18 seats, C: 21 seats, D: 21 seats, E: 78 seats, F: 84 seats, Modified divisor 4580

5.1.11.15.

a. A: 4 seats, B: 4 seats, C: 2 seats

b. It is not possible to assign 11 seats with Hamilton's method in this case.

c. States A and C are the same size, so they have the same decimal value. They would both get an additional seat at the same time but there is only one seat to give. Answers will vary on fair solutions.

d. Yes, with a modified divisor of 1200 we get A: 5 seats, B: 5 seats and C: 1 seat.

5.1.11.16.

a. A: 5 seats, B: 5 seats, C: 0 seats. Each state must have at least one seat so that rule is not met.

b. With 11 seats we have 5 seats, B: 5 seats, C: 1 seat, so it works in this case.

c. Answers will vary on fair solutions.

5.1.11.17.

2010: Douglass: 8, Parks: 4, King: 10, Du Bois: 11, Lewis: 17

2020: Douglass: 8, Parks: 5, King: 10, Du Bois: 11, Lewis: 16

The populations in King and Lewis counties grew, but Parks got an extra seat from Lewis. This doesn't seem fair.

5.1.11.18.

2010: Gray: 10, Castile: 12, Brown: 4, Taylor: 21, Floyd: 15

2020: Gray: 11, Castile: 12, Brown: 4, Taylor: 20, Floyd: 15

The populations in Gray and Floyd counties grew, and Gray got an extra seat from Taylor. This seems reasonable because Gray grew more than Floyd did.

5.1.11.19.

a. Clatsop: 2 counselors, Siletz: 11 counselors

b. The divisor was 698.38, so 4 new guidance counselors should be hired for Cayuse.

c. Clatsop: 3 counselors, Siletz: 10 counselors, Cayuse: 4 counselors

d. Cayuse did get 4 counselors, but one of the counselors from Siletz went to Clatsop. That doesn't seem fair because their populations didn't change.

5.1.11.20.

a. Tubman: 6 art teachers, Blackshear: 2 art teachers

b. The divisor was 78.375 so 2 new art teachers should be hired for Banneker.

c. Tubman: 5 art teachers, Blackshear: 2 art teachers, Banneker: 3 art teachers

d. In the new apportionment, Banneker got 3 art teachers, taking one from Tubman school. The quota for Banneker was 2.947, so it seems reasonable that Banneker would get 3 art teachers.

5.2 · Voting Methods

5.2.12 · Exercises

5.2.12.1.

Number of voters	3	3	1	3	2
1st choice	A	A	B	B	C
2nd choice	B	C	A	C	A
3rd choice	C	B	C	A	B

5.2.12.2.

Number of voters	2	2	2	3	3
1st choice	A	B	B	C	C
2nd choice	B	A	C	A	B
3rd choice	C	C	A	B	A

5.2.12.3.

a. There are 47 voters.

b. A majority is 24 votes.

c. Atlanta wins the plurality method with 19 votes.

d. Buffalo wins the Instant Runoff Method with 28 votes.

e. The points are: Atlanta 94, Buffalo 111 and Chicago 77. Buffalo wins the Borda Count Method.

f. The points are: Buffalo 2, Atlanta 1. Buffalo wins with Copeland's method.

5.2.12.4.

a. There are 33 voters.

b. A majority is 17 votes.

c. Abdulla wins the plurality method with 14 votes.

d. Abdulla wins the Instant Runoff Method with 18 votes.

e. The points are: Abdulla 65, Beck 61 and Cantos 72. Cantos wins the Borda Count Method.

f. The points are: Abdulla 1, Cantos 2. Cantos wins with Copeland's method.

5.2.12.5.

a. There are 12 voters.

b. A majority is 7 votes.

c. Biology wins the plurality method with 5 votes.

d. Biology wins the Instant Runoff Method with 7 votes.

e. The points are: Art 22, Biology 26 and Calculus 24. Biology wins the Borda Count Method.

f. The points are: Biology 2, Calculus 1. Biology wins with Copeland's method.

5.2.12.6.

a. There are 26 voters.

b. A majority is 14 votes.

c. California wins the plurality method with 11 votes.

d. There is a tie between Barbados and California in the Instant Runoff Method with 13 votes each.

e. The points are: Alaska 49, Barbados 54 and California 53. Barbados wins the Borda Count Method.

f. The points are: Barbados 1, California 2. California wins with Copeland's method.

5.2.12.7.

a. There are 31 votes.

b. A majority is 16 votes.

c. F wins the plurality method with 12 votes.

d. E wins the Instant Runoff Method with 16 votes.

e. The points are: D 66, E 59 and F 61. D wins the Borda Count Method.

f. The points are: D 2, E 1. D wins with Copeland's method.

5.2.12.8.

a. There are 51 voters.

b. A majority is 26 votes.

c. G wins the plurality method with 18 votes.

d. H wins the Instant Runoff Method with 26 votes.

e. The points are: G 97, H 106 and I 103. H wins the Borda Count Method.

f. The points are: H 2, I 1. H wins with Copeland's method.

5.2.12.9.

a. There are 460 voters.

b. A majority is 231 votes.

c. A wins the plurality method with 150 votes.

d. A wins the Instant Runoff Method with 290 votes.

e. The points are: A 1140, B 1060, C 1160 and D 1240. D wins the Borda Count method.

f. The points are: A 1, B 1, C 2, D 2. C and D tie with Copeland's method.

5.2.12.10.

a. There are 70 voters.

b. A majority is 36 votes.

c. E wins the plurality method with 27 votes.

d. F wins the Instant Runoff Method with 40 votes.

e. The points are: E 180, F 187, G 166 and H 167. F wins the Borda Count Method.

f. The points are: E 1.5, F 2, G 1, H 1.5. F wins with Copeland's method.

5.2.12.11.

a. There are 92 voters.

b. A majority is 47 votes.

c. K wins the plurality method with 38 votes.

d. K wins the Instant Runoff Method with 54 votes.

e. The points are: I 273, J 152, K 267 and L 228. I wins the Borda Count Method.

f. The points are: I 2, K 3, L2. K wins with Copeland's method.

5.2.12.12.

a. There are 157 voters.

b. A majority is 79 votes.

c. M wins the plurality method with 70 votes.

d. O wins the Instant Runoff Method with 87 votes.

e. The points are: M 472, N 345, O 420, P 331. M wins the Borda Count Method.

f. The points are: M 2, N 1, O 2, P 1. M and O tie with Copeland's method.

5.2.12.13.

a. There are 90 voters.

b. A majority is 46 votes.

c. Q wins the plurality method with 26 votes.

d. S wins the Instant Runoff Method with 50 votes.

e. The points are: Q 250, R 201, S 243 and T 206. Q wins the Borda Count Method.

f. The points are: Q 2, R 1 and S 3. S wins with Copeland's method.

5.2.12.14.

a. There are 110 voters.

b. A majorty is 56 votes.

c. W wins the plurality method with 30 votes.

d. X wins the Instant Runoff Method with 66 votes.

e. The points are: U 262, V 285, w 255 and X 298. X wins the Borda Count Method.

f. The points are: U 1, V 1, W 1 and X 3. X wins with Copeland's method.

5.2.12.15.

a. There are 107 voters.

b. A majority is 54 votes.

c. E wins the plurality method with 39 votes.

d. B wins the Instant Runoff Method with 54 votes.

e. The points are: A 357, B 398, C 305, D 219, E 326. B wins the Borda Count Method.

f. The points are: A 2, B 4, C 2, D 1, E 1. B wins with Copeland's method.

5.2.12.16.

a. There are 92 voters.

b. A majority is 47 votes.

c. F and G tie in the plurality method with 28 votes each.

d. I wins the Instant Runoff Method with 47 votes.

e. The points are: F 324, G 296, H 236, I 296, J 228. F wins the Borda Count Method.

f. The points are: F 3, G 2.5, I 3.5, J 1. I wins with Copeland's method.

5.2.12.17.

a. There are 127 voters.

b. A majority is 64 votes.

c. K wins the plurality method with 35 votes.

d. M wins the Instant Runoff Method with 79 votes.

e. The points are: K 430, L 402, M 376, N 375, O 322. K wins the Borda Count Method.

f. The points are: K 4, L 2, M 2, N 2. K wins with Copeland's method.

5.2.12.18.

a. There are 198 voters.

b. A majority is 100 votes.

c. Q wins the plurality method with 52 votes.

d. Q and R tie in the Instant Runoff Method with 99 votes each.

e. The points are: P 408, Q 606, R 654, S 693, T 609. S wins the Borda Count Method.

f. The points are: Q 2.5, R 2.5, S 3, T 2. S wins with Copeland's method.

5.3 · The Popular Vote, Electoral College and Electoral Power

5.3.4 · Exercises

5.3.4.1.

The president is elected through a process called The Electoral College. States send a certain number of electors, based on their populations, and whichever candidate gets the most votes from these electors becomes president.

5.3.4.2.

There are 538 electors in the Electoral College.

5.3.4.3.

Two senators represent each state.

5.3.4.4.

The term of a U.S. Senator is 6 years.

5.3.4.5.

There are 435 Representatives in the U.S. House of Representatives.

5.3.4.6.

The term of a U.S. Representative is 2 years.

5.3.4.7.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Gandhi	450,000	9	2	11
Mandela	150,000	3	2	5
Gbowee	600,000	12	2	14
Total	1,200,000	24	6	30

This state has 30 electors.

b. A majority of electoral votes would be 16 votes.

5.3.4.8.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Johnson	225,000	3	2	5
Rivera	450,000	6	2	8
Milk	750,000	10	2	12
Total	1,425,000	19	6	25

This state has 25 electors.

b. A majority of electoral votes would be 13 votes.

5.3.4.9.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Tamez	280,000	7	2	9
Teters	200,000	5	2	7
Herrington	400,000	10	2	12
Osawa	360,000	9	2	11
Total	1,240,000	31	8	39

This state has 39 electors.

b. A majority of electoral votes would be 20 electoral votes.

5.3.4.10.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Johnson	120,000	2	2	4
Jackson	480,000	8	2	10
Cox	720,000	12	2	14
Browne	420,000	7	2	9
Total	1,740,000	29	8	37

This state has 37 electors.

b. A majority of electoral votes would be 19 electoral votes.

5.3.4.11.

a.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Gandhi	216,000	234,000	0	11
Mandela	37,500	112,500	0	5
Gbowee	489,450	110,550	14	0
Total Votes	742,950	457,050	14	16

A wins the popular vote with 61.9% of the votes.

b. B wins the electoral college and becomes the president with 53.5% of the electoral votes.

5.3.4.12.

a.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Johnson	115,650	109,350	5	0
Rivera	235,800	214,200	8	0
Milk	117,750	632,250	0	12
Total	469,200	955,800	13	12

B wins the popular vote with 67.1% of the vote.

b. A wins the electoral college and becomes the president with 52% of the electoral votes.

5.3.4.13.

a.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Tamez	95,480	184,250	0	9
Teters	104,200	95,800	7	0
Herrington	203,600	196,400	12	0
Osawa	46,080	313,920	0	11
Total Votes	449,360	790,640	19	20

B wins the popular vote with 63.6% of the vote.

b. B wins the electoral college and becomes the president with 51% of the electoral votes.

5.3.4.14.

a.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Johnson	52,440	67,560	0	4
Jackson	418,080	61,920	10	0
Cox	319,680	400,320	0	14
Browne	350,280	69,270	9	0
Total Votes	1,140,480	599,520	19	18

A wins the popular vote with 65.5% of the vote.

b. A wins the electoral college and becomes the president with 51.4% of the electoral votes.

5.3.4.15.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 50,000 people
Gandhi	450,000	9	2	11	1.22
Mandela	150,000	3	2	5	1.67
Gbowee	600,000	12	2	14	1.17

The state of Mandela has the most electoral power.

b. The state of Gbowee has the least electoral power.

5.3.4.16.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 50,000 people
Johnson	225,000	3	2	5	1.67
Rivera	450,000	6	2	8	1.33
Milk	750,000	10	2	12	1.2

The state of Johnson has the most electoral power.

b. The state of Milk has the least electoral power.

5.3.4.17.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 50,000 people
Tamez	280,000	7	2	9	1.29
Teters	200,000	5	2	7	1.40
Herrington	400,000	10	2	12	1.20
Osawa	360,000	9	2	11	1.22

The state of Teters has the most electoral power.

b. The state of Herrington has the least electoral power.

5.3.4.18.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 50,000 people
Johnson	120,000	2	2	4	2.00
Jackson	480,000	8	2	10	1.25
Cox	720,000	12	2	14	1.17
Browne	420,000	7	2	9	1.29

The state of Johnson has the most electoral power.

b. The state of Cox has the least electoral power.

5.3.4.19.

Possible combinations are Gandhi/Mandela/Gbowee, Gandhi/Mandela, Mandela/Gbowee, or Gandhi/Gbowee. The minimum number of votes needed is 300,002.

5.3.4.20.

Possible combinations are Johnson/Rivera/Milk, Johnson/Rivera, Rivera/Milk, or Johnson/Milk. The minimum number of votes needed is 337,502.

5.3.4.21.

Possible combinations are Tamez/Teters/Herrington/Osawa, Tamez/Teters/Herrington, Tamez/Teters/Osawa, Tamez/Herrington/Osawa, Teters/Herrington/Osawa, Tamez/Herrington, Tamez/Osawa, or Herrington/Osawa. The minimum number of votes needed is 320,002.

5.3.4.22.

Possible combinations are Johnson/Jackson/Cox/Browne, Johnson/Jackson/Cox, Johnson/Jackson/Browne, Johnson/Cox/Browne, Jackson/Cox/Browne, Jackson/Cox, Jackson/Browne, or Cox/Browne. The minimum number of votes needed is 450,002.

5.4 · Gerrymandering and How to Measure It

5.4.5 · Exercises

5.4.5.1.

Redistricting happens every 10 years after the census is completed. A state might not change its districts unless the number of seats in the U.S. House of Representatives has changed.

5.4.5.2.

Each state government is in charge of drawing the lines.

5.4.5.3.

The rules are districts must be contiguous and you can't gerrymander by race.

5.4.5.4.

The two ways to gerrymander are by packing, putting all of one party in a district, or by cracking, splitting up a party into many districts, so their votes are ineffective.

5.4.5.5.

The most proportional representation would be 4 Republican seats and 3 Democratic seats.

5.4.5.6.

The most proportional representation would be 5 Republican seats and 7 Democratic seats.

5.4.5.7.

The most proportional representation would be 2 Republican seats and 1 Democratic seat.

5.4.5.8.

The most proportional representation would be 3 Republican seats and 6 Democratic seats.

5.4.5.9.

The most proportional representation would be 6 Republican seats, 4 Democratic seats and 1 Green Party seat.

5.4.5.10.

The most proportional representation would be 6 Republican seats, 10 Democratic seats and 2 Progressive Party seats.

5.4.5.11.

a. A majority is 4 votes.

b. The Democrats won 2 seats and the Republicans won 2 seats.

c. The efficiency gap is $6/28 = 21.43\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	1	6	1	$6-4=2$
2	3	4	3	$4-4=0$
3	6	1	$6-4=2$	1
4	7	0	$7-4=3$	0
Total	17	11	9	3

d. Each seat is worth 25% of the voters.

e. The efficiency gap is worth less than one seat (0.86 seats).

f. This map is ok because the efficiency gap is less than one seat. It should either be D3, R1 or D2, R2.

5.4.5.12.

a. A majority is 5 votes.

b. The Democrats won all 4 seats.

c. The efficiency gap is $14/36 = 38.89\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	5	4	$5-5=0$	4
2	5	4	$5-5=0$	4
3	6	3	$6-5=1$	3
4	5	4	$5-5=0$	4
Total	21	15	1	15

d. Each seat is worth 25% of the voters.

e. The efficiency gap is worth 1.56 seats.

f. This map is not fair because the efficiency gap is more than one seat. A more fair map would be D3, R1.

5.4.5.13.

a. A majority is 3 votes.

b. The Democrats won 1 seat and the Republicans won 4 seats.

c. The efficiency gap is $8/25 = 32\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	2	3	2	$3-3=0$
2	2	3	2	$3-3=0$
3	2	3	2	$3-3=0$
4	4	1	$4-3=1$	1
5	2	3	2	$3-3=0$
Total	12	13	9	1

d. Each seat is worth 20% of the voters.

e. The efficiency gap is worth 1.6 seats.

f. This map is not fair because the efficiency gap is more than one seat. A fairer map would be D3, R2 or D2, R3.

5.4.5.14.

a. A majority is 4 votes.

b. The Democrats won 3 seats and the Republicans won 2 seats.

c. The efficiency gap is $3/35 = 8.57\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	3	4	3	$4-4=0$
2	5	2	$5-4=1$	2
3	3	4	3	$4-4=0$
4	5	2	$5-4=1$	2
5	5	2	$5-4=1$	2
Total	21	14	9	6

d. Each seat is worth 20% of the voters.

e. The efficiency gap is worth less than one seat (0.43).

f. This map is fair because the efficiency gap is around 8% and the representation is exactly proportional to the population.

5.4.5.15.

a. A majority is 5 votes.

b. The Democrats won 2 seats and the Republicans won 3 seats.

c. The efficiency gap is $4/45 = 8.89\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	2	7	2	$7-5=2$
2	5	4	$5-5=0$	4
3	2	7	2	$7-5=2$
4	5	4	$5-5=0$	4
5	4	5	4	$5-5=0$
Total	18	27	8	12

d. Each seat is worth 20% of the voters.

e. The efficiency gap is worth less than one seat (0.44).

f. This map is fair because the efficiency gap is around 8% and the representation is exactly proportional to the population.

5.4.5.16.

a. A majority is 6 votes.

b. The Democrats won 1 seats and the Republicans won 4 seats.

c. The efficiency gap is $17/55 = 30.91\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	5	6	5	$6-6=0$
2	5	6	5	$6-6=0$
3	7	4	$7-6=1$	4
4	5	6	5	$6-6=0$
5	5	6	5	$6-6=0$
Total	27	28	21	4

d. Each seat is worth 20% of the voters.

e. The efficiency gap is worth 1.55 seats.

f. This map is not fair because the efficiency gap is more than one seat. A more fair map would be D2, R3.

5.4.5.17.

a. A majority is 3 votes.

b. The Democrats won 5 seats and the Republicans won 1 seat.

c. The efficiency gap is $12/30 = 40\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	3	2	$3-3=0$	2
2	3	2	$3-3=0$	2
3	3	2	$3-3=0$	2
4	3	2	$3-3=0$	2
5	3	2	$3-3=0$	2
6	0	5	0	$5-3=2$
Total	15	15	0	12

d. Each seat is worth 16.67% of the voters.

e. The efficiency gap is worth 2.4 seats.

f. This map is not fair because the efficiency gap is more than one seat. A more fair map would be D3, R3 because the population is evenly split.

5.4.5.18.

a. A majority is 4 votes.

b. The Democrats won 1 seat and the Republicans won 5 seats.

c. The efficiency gap is $16/42 = 38.10\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	3	4	3	$4-4=0$
2	2	5	2	$5-4=1$
3	3	4	3	$4-4=0$
4	3	4	3	$4-4=0$
5	7	0	$7-4=3$	0
6	3	4	3	$4-4=0$
Total	21	21	17	1

d. Each seat is worth 16.67% of the voters.

e. The efficiency gap is worth 2.3 seats.

f. This map is not fair because the efficiency gap is more than one seat. A more fair map would be D3, R3 because the population is evenly split.

5.4.5.19.

a. A majority is 5 votes.

b. The Democrats won 1 seat and the Republicans won 5 seats.

c. The efficiency gap is $6/54 = 11.11\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	2	7	2	$7-5=2$
2	2	7	2	$7-5=2$
3	9	0	$9-5=4$	0
4	2	7	2	$7-5=2$
5	1	8	1	$8-5=3$
6	4	5	4	$5-5=0$
Total	20	34	15	9

d. Each seat is worth 16.67% of the voters.

e. The efficiency gap is worth less than 1 seat. (0.67)

f. This map is not fair because even though the efficiency gap is less than one seat, The Democrats should have at least 2 seats. A more fair map would be D2, R4.

5.4.5.20.

a. A majority is 6 votes.

b. The Democrats won 4 seats and the Republicans won 2 seats.

c. The efficiency gap is $26/66 = 39.39\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	0	11	0	$11-6=5$
2	8	3	$8-6=2$	3
3	0	11	0	$11-6=5$
4	6	5	$6-6=0$	5
5	6	5	$6-6=0$	5
6	6	5	$6-6=0$	5
Total	26	40	2	28

d. Each seat is worth 16.67% of the voters.

e. The efficiency gap is worth 2.36 seats.

f. This map is not fair because the efficiency gap is worth more than one seat. The Republicans should have at least 3 seats. A more fair map would be D3, R3 or D2, R4.

5.4.5.21.

a. The most seats that can be won by the Democrats is 4 seats.

b. The most seats that can be won by the Republicans is 4 seats.

5.4.5.22.

a. The most seats that can be won by the Democrats is 4 seats.

b. The most seats that can be won by the Republicans is 4 seats.

5.4.5.23.

a. The most seats that can be won by the Democrats is 4 seats.

b. The most seats that can be won by the Republicans is 5 seats.

5.4.5.24.

a. The most seats that can be won by the Democrats is 6 seats.

b. The most seats that can be won by the Republicans is 4 seats.

5.5 · Chapter 5 Review

· Review Exercises

5.5.1.

a. A: 10 seats, B: 17 seats, C: 12 seats, D: 39 seats

b. A: 10 seats, B: 16 seats, C: 12 seats, D: 40 seats, Modified divisor 9,600

c. A: 10 seats, B: 17 seats, C: 12 seats, D: 39 seats, Divisor 9,793.6

d. A: 10 seats, B: 17 seats, C: 12 seats, D: 39 seats, Divisor 9,793.6

5.5.2.

a. A: 12 seats, B: 19 seats, C: 14 seats, D: 45 seats

b. A: 11 seats, B: 19 seats, C: 14 seats, D: 46 seats, Modified divisor 8,300

c. A: 11 seats, B: 19 seats, C: 14 seats, D: 46 seats, Modified divisor 8,400

d. A: 11 seats, B: 19 seats, C: 14 seats, D: 46 seats, Modified divisor 8,400

5.5.3.

a. A: 46 seats, B: 18 seats, C: 7 seats, D: 21 seats, E: 8 seats

b. A: 47 seats, B: 18 seats, C: 7 seats, D: 20 seats, E: 8 seats, Modified divisor 12,400

c. A: 47 seats, B: 18 seats, C: 7 seats, D: 20 seats, E: 8 seats, Modified divisor 12,555

d. A: 47 seats, B: 18 seats, C: 7 seats, D: 20 seats, E: 8 seats, Modified divisor 12,555

5.5.4.

a. A: 58 seats, B: 22 seats, C: 9 seats, D: 26 seats, E: 10 seats

b. A: 59 seats, B: 22 seats, C: 8 seats, D: 26 seats, E: 10 seats, Modified divisor 9,890

c. A: 58 seats, B: 22 seats, C: 9 seats, D: 26 seats, E: 10 seats, Divisor 10,085.6

d. A: 58 seats, B: 22 seats, C: 9 seats, D: 26 seats, E: 10 seats, Divisor 10,085.6

5.5.5.

a. There are 134 voters.

b. A majority is 68 votes.

c. H wins the plurality method with 35 votes.

d. F wins the Instant Runoff Method with 71 votes.

e. The points are: E 275, F 364, G 372 and H 329. G wins in the Borda count method.

f. The points are: F 3, G 2, H 1. F wins with Copeland's method.

5.5.6.

a. There are 117 voters.

b. A majority is 59 votes.

c. K wins the plurality method with 38 votes.

d. L wins the Instant Runoff Method with 66 votes.

e. The points are: I 313, J 304, K 261, L 292. I wins in the Borda count method.

f. The points are: I 2, J 2, L 2. I, J and L tie with Copeland's method.

5.5.7.

a. There are 166 voters.

b. A majority is 84 votes.

c. N wins the plurality method with 37 votes.

d. P wins the Instant Runoff Method with 101 votes.

e. The points are: M 486, N 482, O 508, P 590, Q 424. P wins the Borda count method.

f. The points are: M 1, N 2, O 3, P 4. P wins with Copeland's method.

5.5.8.

a. There are 221 voters.

b. A majority is 111 votes.

c. U wins the plurality method with 47 votes.

d. T wins the Instant Runoff Method with 138 votes.

e. The points are: R 495, S 705, T 768, U 642, V 705. T wins the Borda count method.

f. The points are: S 3, T 4, U 2, V 1. T wins with Copeland's method.

5.5.9.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Fonville	825,000	15	2	17
Gurley	550,000	10	2	12
Nevarez	275,000	5	2	7
Total	1,650,000	30	6	36

This state has 36 electors.

b. A majority of electoral votes would be 19 votes.

5.5.10.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors
Arbery	720,000	12	2	14
Monterrosa	360,000	6	2	8
Bland	240,000	4	2	6
Davis	480,000	8	2	10
Total	1,800,000	30	8	38

This state has 38 electors.

b. A majority of electoral votes would be 20 votes.

5.5.11.

a.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Fonville	684,750	140,250	17	0
Gurley	257,400	292,600	0	12
Nevarez	132,275	142,725	0	7
Total Votes	1,074,425	575,575	17	19

A wins the popular vote with 65.1% of the votes.

b. B wins the electoral college and becomes the president with 52.8% of the electoral votes.

5.5.12.

a.

State	Votes for Candidate A	Votes for Candidate B	Number of Electoral Votes for A	Number of Electoral Votes for B
Arbery	372,240	347,760	14	0
Monterrosa	38,880	321,120	0	8
Bland	134,640	105,360	6	0
Davis	104,160	375,840	0	10
Total	649,920	1,150,080	20	18

B wins the popular vote with 63.9% of the vote.

b. A wins the electoral college and becomes the president with 52.6% of the electoral votes.

5.5.13.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 55,000 people
Fonville	825,000	15	2	17	1.13
Gurley	550,000	10	2	12	1.20
Nevarez	275,000	5	2	7	1.40

The state of Nevarez has the most electoral power.

b. The state of Fonville has the least electoral power.

5.5.14.

a.

State	Population	Number of Representatives	Number of Senators	Number of Electors	Electoral Votes per 60,000 people
Arbery	720,000	12	2	14	1.17
Monterrosa	360,000	6	2	8	1.33
Bland	240,000	4	2	6	1.50
Davis	480,000	8	2	10	1.25

The state of Bland has the most electoral power.

b. The state of Arbery has the least electoral power.

5.5.15.

a. A majority is 5 votes.

b. The Democrats won 1 seat and the Republicans won 4 seats.

c. The efficiency gap is $20/45 = 44.4\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	4	***5***	4	$5-5=0$
2	4	***5***	4	$5-5=0$
3	***9***	0	$9-5=4$	0
4	4	***5***	4	$5-5=0$
5	4	***5***	4	$5-5=0$
Total	25	20	20	0

d. Each seat is worth 20% of the voters.

e. The efficiency gap is worth 2.2 seats.

f. This map is not fair because the efficiency gap is more than two seats. A more fair map would be D3, R2.

5.5.16.

a. A majority is 4 votes.

b. The Democrats won 5 seats and the Republicans won 1 seat.

c. The efficiency gap is $16/42 = 38.10\%$

District	D Votes	R Votes	D Surplus Votes	R Surplus Vote
1	*4*	3	$4-4=0$	3
2	*4*	3	$4-4=0$	3
3	*4*	3	$4-4=0$	3
4	*4*	3	$4-4=0$	3
5	*4*	3	$4-4=0$	3
6	1	***6***	1	$6-4=2$
Total	21	21	1	17

d. Each seat is worth 16.67% of the voters.

e. The efficiency gap is worth 2.3 seats.

f. This map is not fair because the efficiency gap is more than two seats. A more fair map would be D3, R3 because the population is evenly split.

5.6 · Federal Budget, Deficit and National Debt

5.6.8 · Exercises

5.6.8.1.

Federal income comes from individual income taxes and business income taxes.

5.6.8.2.

The two types of federal spending are mandatory and discretional spending.

5.6.8.3.

Twelve appropriations bills must be passed to approve the new federal budget.

5.6.8.4.

The deadline for the new budget is September 30th each year.

5.6.8.5.

A deficit is a shortfall in a single year and debt is the total of all the money owed.

5.6.8.6.

The Gross Domestic Product is the total of all the finished goods and services produced by a country in a specific period of time.

5.6.8.7.

a 4.3 billion

b 12.567 trillion

c 500 million or 0.5 billion

d 6.04 million

5.6.8.8.

a 63.651 trillion

b 93.6 million

c 119.93 billion or 0.11993 trillion

d 6.001 billion

5.6.8.9.

a 5,700,000

b 9,220,000,000,000

c 100,200,000,000

d 250,000,000,000

5.6.8.10.

a 520,000,000,000,000

b 1,490,000,000

c 9,070,000,000,000

d 800,000,000

5.6.8.11.

a The debt to GDP ratio for Columbia is approximately 55.13%.

b The amount of debt owed per person in Columbia is approximately C$ 5.541 million.

c The amount of interest paid per year per person in Columbia is approximately C$ 0.34 million or C$ 340,040.

5.6.8.12.

a The debt to GDP ratio for Pakistan is approximately 88.67%.

b The amount of debt owed per person in Pakistan is approximately Rs 0.115 million or Rs 115,655.

c The amount of interest paid per year per person in Pakistan is approximately Rs 9,695.

5.6.8.13.

a The debt to GDP ratio for Poland is approximately 61.74%.

b The amount of debt owed per person in Poland is approximately 31,773 zł.

c The amount of interest paid per year per person in Pakistan is approximately 1.36 zł.

5.6.8.14.

a The debt to GDP ratio for Australia is approximately 33.49%.

b The amount of debt owed per person in Australia is approximately A$ 26,115.

c The amount of interest paid per year per person in Australia is approximately A$ 798.

5.6.8.15.

a The debt to GDP ratio for South Africa is approximately 60.39%.

b The amount of debt owed per person in South Africa is approximately R. 68,550.

c The amount of interest paid per year per person in South Africa is approximately R. 3,612.

5.6.8.16.

a The debt to GDP ratio for Malaysia is approximately 54.88%.

b The amount of debt owed per person in Malaysia is approximately RM 14,541.

c The amount of interest paid per year per person in Malaysia is approximately RM 508.

5.6.8.17.

a Approximately \$7.668 billion is budgeted for transportation.

b Approximately \$2,255 is budgeted for health and human resources per person.

c Approximately \$3.834 billion is budgeted for public safety in total.

5.6.8.18.

a Approximately \$3.115 billion was budgeted for K-12 education.

b Approximately \$77 was budgeted for local governements per person.

c Approximately \$623 million was budgeted for corrections in total.

References

About the electors. (2019, August 27). National Archives. https://www.archives.gov/electoral-college/electors

American Civil Liberties Union. (2020). *Voting Rights Act: Major Dates in History.* American Civil Liberties Union. https://www.aclu.org/voting-rights-act-major-dates-history

Ballotpedia Staff. (n.d.). *Presidential Election, 2016.* Ballotpedia. Retrieved 2020, from https://ballotpedia.org/Presidential_election,_2016

Best, J. (2004). Presidential selection: Complex problems and simple solutions. *Political Science Quarterly, 119*(1), 39–59.

Brown, S. (2015, March 22). *Baltimore Professor Working To Make Higher Mathematics Accessible To All.* CBS Baltimore. https://baltimore.cbslocal.com/2015/03/22/baltimore-professor-working-to-make-higher-mathematics-accessible-to-all/

Chichilnisky, G. (1980). Social choice and the topology of spaces of preferences. *Advances in Mathematics, 37*(2), 165–176. https://doi.org/10.1016/0001-8708(80)90032-8

Chichilnisky, G. (2018). *Graciela Chichilnisky – Graciela Chichilnisky.* Bio. https://chichilnisky.com/graciela-chichilnisky/

Clayton, D. (2015). The electoral college: An idea whose time has come and gone. *The Black Scholar, 37*(3), 28–41.

David Harold Blackwell: National Visionary Leadership Project: African American History. (2013). http://www.visionaryproject.org/blackwelldavid/

Farley, J. (2020). *Jonathan Farley.* www.Latticetheory.Net. http://www.latticetheory.net/

Freidenreich, H. (2020). *Hilda Geiringer.* Jewish Women's Archive. https://jwa.org/encyclopedia/article/geiringer-hilda

Giridharadas, A., and Bradsher, K. (2006, October 13). Microloan Pioneer and His Bank Win Nobel Peace Prize. *The New York Times.* https://www.nytimes.com/2006/10/13/business/14nobelcnd.html

Hermoza, L. C. (2015, October 5). *Associate Professor In Baltimore Uses Mathematics To Counter Terrorism.* CBS Baltimore. https://baltimore.cbslocal.com/2015/10/05/associate-professor-in-baltimore-uses-mathematics-to-counter-terrorism/

History.com Editors. (2018a, September 14). *Native American Cultures.* HISTORY. https://www.history.com/topics/native-american-history/native-american-cultures

History.com Editors. (2018b, October 4). *Compromise of 1850.* HISTORY. https://www.history.com/topics/abolitionist-movement/compromise-of-1850

How Slavery Helped Build a World Economy. (2003, January 3). Nationalgeographic.Com. https://www.nationalgeographic.com/news/2003/1/how-slavery-helped-build-a-world-economy/#close

Indian Land Trust Tenure Foundation. (2020). *History – ILTF.* Indian Land Tenure Foundation. https://iltf.org/land-issues/history/

Jonathan Farley, Mathematics – Martin Luther King Jr. Scholars. (2005). MLK Visiting Professors and Scholars Program. https://mlkscholars.mit.edu/jfarley/

Manning, B. (2019, November 5). *Overdue recognition.* Wheaton College Massachusetts. https://wheatoncollege.edu/news/overdue-recognition/

McDonald, M. P. (2016). 2016 *November General Election Turnout Rates.* Electproject.Org. http://www.electproject.org/2016g

McNeill, L. (2019). *The woman who reshaped maths.* BBC.com; BBC Future. https://www.bbc.com/future/article/20191031-hilda-geiringer-mathematician-who-fled-the-nazis

Paiva, J. M. (2011). On introducing proportionality in american presidential elections: An historical analysis. *The Political Quarterly, 82*(3), 435–447.

Ranked Choice Voting Resource Center. (2020). Ranked Choice Voting Resource Center. https://www.rankedchoicevoting.org/where_used

Renton, P. (2008, December 5). *A Brief History of Microlending.* Lend Academy. https://www.lendacademy.com/a-brief-history-of-microlending/

Robertson, E., and O'Connor, J. (2020). *Hilda Geiringer - Biography.* MacTutor History of Mathematicians Archive. https://mathshistory.st-andrews.ac.uk/Biographies/Geiringer/

Siegle, L. (2010, November 14). Graciela Chichilnisky's innovation: carbon capturing. *The Guardian.* https://www.theguardian.com/environment/2010/nov/14/graciela-chichilnisky-carbon-capture-global-thermostat

Stafford, J. B. (2006, March 3). *Does 1+1=terror? Using math models to calculate risks.* Stanford University. https://news.stanford.edu/news/2006/march8/mterror-030806.html

Wikipedia Contributors. (2019, January 27). *Muhammad Yunus.* Wikipedia; Wikimedia Foundation. https://en.wikipedia.org/wiki/Muhammad_Yunus

Williams, S. W. (2008). *Who are the greatest Black Mathematicians?* www.Math.Buffalo.edu. http://www.math.buffalo.edu/mad/PEEPS/blackwell_david.html

Worrall, S. (2018, June 9). *When, How Did the First Americans Arrive? It's Complicated.* Nationalgeographic.com. https://www.nationalgeographic.com/news/2018/06/when-and-how-did-the-first-americans-arrive--its-complicated-/