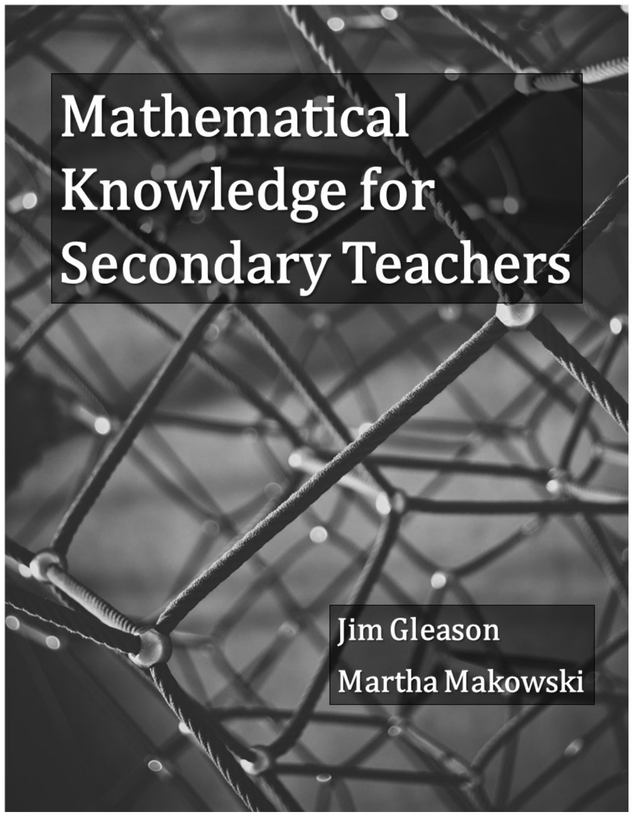
Mathematical Knowledge for Secondary Teachers

Jim Gleason and Martha Makowski

Contents


Mathematical
Knowledge for
Secondary Teachers
Jim Gleason
Martha Makowski

Preface

The knowledge possessed by an excellent secondary mathematics teacher draws on a variety of skills, beliefs, and passions related to the practice of teaching. A large portion of this knowledge is acquired through experience working with students. However, excellent teachers develop certain aspects of knowledge through structured learning environments like a textbook or university course.

This textbook is inspired by Usiskin et al. [2003] and designed to help current and future teachers explore a unique blend of *mathematical content knowledge* used in the practice of teaching mathematics at the secondary and early post-secondary level. Content examining how K-12 students learn mathematics or on classroom management techniques related to mathematical teaching are left to other texts. There are many excellent resources for this pedagogical content knowledge, particularly from the National Council of Teachers of Mathematics (NCTM) and the Association of Mathematics Teacher Educators (AMTE).

DESCRIBE CHOICE OF COVER ART.

For the Student

For the Instructor

(NOTE THE DIFFERENCE IN CHAPTER LENGTHS)

One Semester Courses

Two Semester Course Sequence

Three Semester Courses

Describe Current process at UA

Focus on Middle School Content

Which Chapters and sections are really most appropriate for Middle School teachers.

Chapter Dependencies

Acknowledgements

APLUS-M project and master teachers

2020 and 2021 Cohorts

Anna Keefe, Ty Campbell, Sara Grace Estill

Part I

FOUNDATIONS

Chapter 1

Mathematics Education Foundations

This first chapter focuses on developing the notion of Mathematical Knowledge for Teaching and examining how content knowledge supports the current standards for mathematical practice and the content standards developed by various educational researchers and mathematics education organizations. Understanding these foundations prior to jumping into the rest of the content provides both an organizational framework and a foundation for the topics covered. Some of these ideas may challenge your ways of thinking, while others reinforce what you already know. We encourage you to go beyond what is written here and to read some of the original research articles and organization standards cited in this chapter.

1.1 History of Mathematics Education in the United States

Discuss the transitions through the different eras. Focus particularly on the space race focus of calculus and the information technology focus on data analysis topics (discrete, probability, statistics).

Equations and covariational reasoning vs. functions?

1.2 Mathematical Knowledge for Teaching

In his presidential address to the American Educational Research Association, Lee Shulman (-Shulman [1986]) popularized the concept of knowledge for teaching that included a specialized content knowledge where "the teacher need not only understand *that* something is so; the teacher must further understand *why* it is so, on what grounds its warrant can be asserted, and under what circumstances our belief in its justification" (p. 9). Researchers in mathematics education further expanded on this idea in the development of domains of Mathematical Knowledge for Teaching. This textbook focuses on the Subject Matter Knowledge side of the Mathematical Knowledge for Teaching.

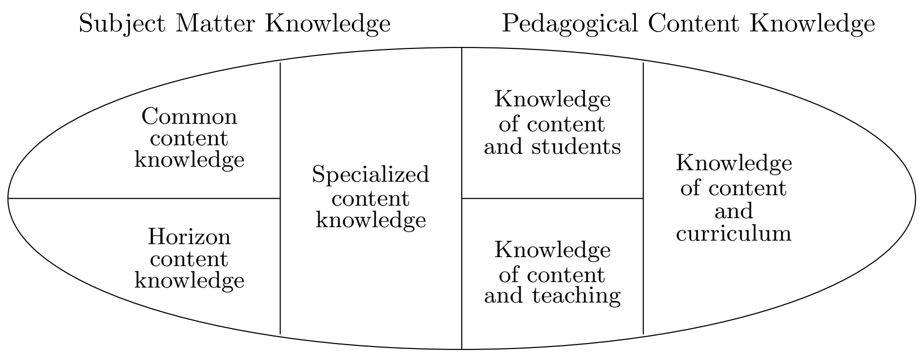


Figure 1.1: Domains of Mathematical Knowledge for Teaching [@Ball2008]

1.2.1 Common Content Knowledge

The foundation for much of the other aspects of mathematical knowledge for teaching is **common content knowledge**, defined as "the mathematical knowledge and skill used in settings other than teaching" [Ball et al., 2008, p. 399]. Common content knowledge provides the individual the ability to solve mathematical problems from the related curriculum and to apply the mathematical knowledge to other fields of knowledge outside of mathematics. Common content knowledge also includes understanding how some of the different mathematical subjects build upon one another, but does not require those studying the material to be able to explain broader patterns. Thus, common content knowledge represents the type of knowledge that we expect secondary and introductory post-secondary students to display.

For instance, common content knowledge related to rational algebraic expressions and functions includes understanding the relationships between rational expressions and rational numbers, how long-division of polynomials relates to long-division of integers, and the properties of the quotient and remainder of a rational expression and its effect on the graph of the related function. While this common content knowledge includes a deeper level of knowledge and a richer web of conceptual understanding by the teacher than is common for most K-12 students, the knowledge still falls within the expectations for the most advanced of these K-12 students. Thus, it is paramount that secondary mathematics teachers develop deep fluency in common content knowledge and the interconnectedness of the many different mathematical concepts in the curriculum. Such a deep and interconnected knowledge base of the teacher provides a requisite foundation in helping K-12 students learn mathematics. However, it is only a first step.

1.2.2 Specialized Content Knowledge

Specialized content knowledge incorporates mathematical knowledge that goes beyond knowledge expected of students, with a focus on the mathematical knowledge that improves the ability of the teacher to assist students learning mathematics [Ball et al., 2008, p. 400]. Thus, specialized content knowledge supplements the common content knowledge that all K-12 students of mathematics need. For example, while common content knowledge for rational algebraic expressions and rational functions include the relationships to the rational numbers, specialized content knowledge could include knowledge of rings and integral domains, thereby improving the teacher's ability to help students to make the connections between various pieces of related content knowledge.

A teacher would also use this specialized content knowledge related to rational expressions when explaining the extent to which two rational expressions are equivalent when common factors of the numerator and denominator cancel. For instance the teacher could explain the ways in which the expressions

$$\frac{(x-1)(x+2)^2}{(x+1)(x+2)} \quad \text{and} \quad \frac{(x-1)(x+2)}{(x+1)}$$

are equivalent and distinct.

1.2.3 Horizon Content Knowledge

Excellent teachers use more than just a knowledge of the content in the current curriculum when teaching students. They also draw on knowledge of what their students have previously learned in mathematics, what mathematics content will be covered in the next few years, and how the current mathematical topic relates to applications outside of mathematics. That is, excellent mathematics teachers see their instruction as part of a continuum, of which their work is only a small part. We call this domain of knowledge **horizon content knowledge**.

For example, a teacher with horizon content knowledge might use students' familiarity with rational numbers to help high school students develop knowledge of rational algebraic expressions and functions. The teacher could also use her knowledge from differential equations to know that the rational functions play a pivotal role in the Laplace transform, helping to determine the amount of time and detail appropriate for teaching about the quotient and remainder theorem for polynomials and how to rewrite rational expressions using partial fractions. The teacher could also use knowledge of the physics curriculum and Boyle's law to help students make connections between rational expressions and functions and other fields of study.

Horizon content knowledge can also allow teachers to help students understand content in other subjects as well. For instance, many of the sciences are becoming more reliant on organizing and understanding data and graphs. Understanding level curves can help students to better use topographical maps. Understanding geometric properties provides artists with tools to create projections in paintings. And venn diagrams and basic set theory is used in many other subjects to organize information and structures.

While each of these pieces of knowledge are not essential for mathematics teachers, the more knowledge one has, the better that teacher can help students learn and apply the critical components of the content.

1.2.4 Exercises

1. Consider the general quadratic expression $ax^2 + bx + c$, where a, b, and c are real numbers such that $a \neq 0$.

   a. Write a list of everything you **know** about the general quadratic expression.

   b. Write a list of everything you can **do** to the general quadratic expression.

   c. How are quadratic expressions different than quadratic equations?

d. How does the factorization of a quadratic expression relate to prime numbers?

e. How much do you know about the ways in which quadratic expressions are used outside of mathematics?

f. Review the mathematics standards for quadratic expressions for your state. How many of the things you have listed for parts (a) and (b) match up with those standards?

g. For each of your answers to a. through e., does the content you have listed or described align most with common content knowledge, specialized content knowledge, or horizon content knowledge?

2. Explain the ways in which the expressions

$$\frac{(x-1)(x+2)^2}{(x+1)(x+2)} \quad \text{and} \quad \frac{(x-1)(x+2)}{(x+1)}$$

are equivalent and distinct.

3. Consider a triangle. A particular high school geometry textbook defines a triangle as "a polygon with three sides." A second textbook defines a triangle as "the figure formed by connecting three non-co-linear points with straight segments." A last textbook defines triangles as "A three-sided figure."

   a) Are all three definitions accurate, or do some allow for shapes that might not be triangles as you understand them to be included within the category of triangle?

   b) In what ways might each definition be considered sloppy? That is, are there any parts of the definition that might not be well-defined? In what ways might each definition make using triangles in future lessons more difficult?

   c) What information do you think each textbook has presented prior to giving its definition for a triangle?

   d) How might horizon content knowledge help an instructor preparing a lesson on triangles decide whether a definition is appropriate or not for her students?

4. In order to better understand the importance of definitions we will consider the word of "square".

   a) Define "square" with as few of words as possible, without using outside resources.

   b) What additional mathematical words would be needed to be defined to understand your definition?

   c) How would your definition of "square" fit within other types of quadrilaterals?

1.3 Mathematical Practice Standards

Just like almost all of the content taught in secondary schools, the development of procedural habits and pieces of information are important parts of a secondary education, the heart of learning lies in developing habits of thinking, perseverance techniques, and developing communication skills to improve the ability to interact with the world around them. For example, the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and the United Nations Office on Drugs and Crime (UNODC) [2019] described cognitive learning outcomes for secondary education such that a student "Knows about local, national, and global governance and accountability systems and structures, understands issues affecting interaction and connectedness of communities at local, national and global levels, (and) develops skills for critical inquiry and analysis" (p. 16). Mathematics, despite it's reputation, aligns well with these goals. Critical thinking, reasoning, and communicating are the most important components of the secondary mathematics curriculum, and are often the most ignored.

The National Council of Teachers of Mathematics [2000] describes these goals in the -Principles and Standards for School Mathematics-, giving them the name 'Process Standards.' The National Governors Association Center for Best Practices and the Council of Chief State School Officers [NGA-CCSSO, 2010] expanded upon

Table 1.1: NCTM Process Standards and Common Core Standards for Mathematical Practice

Standards for Mathematical Practice	Process Standards
Make sense of problems and persevere in solving them.	Problem Solving
Reason abstractly and quantitatively.	Reasoning and Proof
Construct viable arguments and critique the reasoning of others.	Communication
Model with mathematics.	Connections
Use appropriate tools strategically.	Representations
Attend to precision	
Look for and make use of structure.	
Look for and express regularity in repeated reasoning.	

these to create the 'Common Core State Standards Standards for Mathematical Practice.' These practice standards are listed in Table 1.1 and more details about these standards can be found in their corresponding publications.

It is worth noting that these practices are expected of students of mathematics at *all* grade levels. In order to help others develop these practices, we must first develop them in ourselves. Only then will we be in a position to create a learning environment that motivates and enables students to grow in these practices.

To support the users of this text in developing these practices, opportunities to apply them are woven into the text, exercises, and projects. However, as any good teacher knows, the learner needs to be actively engaged in order for an objective to be reached. As you work through the text, pay attention to how the arguments are presented and seek to understand the process behind the mathematical content, rather than just procedures. When completing the exercises and projects, do not just try to get an answer. Instead, take some time to grapple with the ideas, think of better ways to communicate what you do not understand, and seek to understand the deeper connections involved in the task.

For the purpose of this text, we group these practice standards into four categories: mathematical problem solving, modeling with mathematics, communicating mathematically, and understanding mathematical structures. We briefly elaborate on each in the following sections.

1.3.1 Mathematical Problem Solving

Mathematical problem solving is perhaps the most widely cited application of mathematics "in the real world". Although generally acknowledge as important, problem solving is a complex process that is difficult to teach. In his book, *How to Solve It: A new aspect of mathematical method*, George Pólya described four phases of the problem solving process (See Figure 1.2 when approaching mathematical problems [Polya, 1957]. While many others have expanded on this problem solving process, much of the thinking around problem solving still traces back to the four phases that Pólya describes.

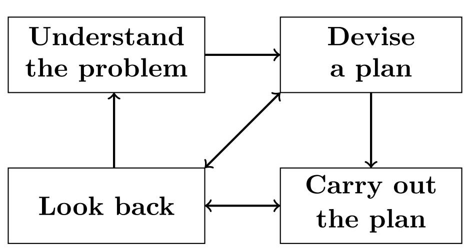


Figure 1.2: Pólya's problem solving process

The first phase is the process of understanding the problem. What this phase looks like varies by problem, but often involves making sure you understand what information is given and relevant. It also includes knowledge of what an answer should look like for the question posed and, where relevant, being able to organize the provided information in a picture or diagram to better understand the situation. So, for example, if the problem is computational, understanding the problem may entail understanding whether positive or negative answers are valid solutions. If the problem is a theorem that needs proving, this phase may involve generating related examples or writing out relevant definitions to make sure you understand the given inputs and implications of the theorem.

The second phase of the problem solving process is to devise a plan. Many students struggle in the problem solving process because they only want to solve problems for which they have previously been given a template. The origin of this desire has roots in the fact that many teachers only assign students problems that are similar to those discussed in class. True mathematical problem solving involves confronting problems for which the solution is not immediately obvious to the problem solver. In problem solving, one may need to connect the current problem to previously solved problems or restate the problem in a different form with an easier solution process. Sometimes a full plan is not possible at the beginning and the solver needs to just plan initial steps and work through those to gather more information about the problem that will help them create a new plan later in the process.

Once an initial plan is devised, it needs to be carried out. This is often the simplest part of the problem solving process, and often the only component that students complete. With today's technology, computer applications can often complete the details of this part of the problem solving process through statistical analyses, computer algebra systems, or graphical programs. While there are benefits of technology, it is important that the students understand the process.

After the plan is carried out, it is important to look back, completing the final phase of the problem solving process. In this step the problem solver determines whether their solution makes sense in the context of the original problem. For example, in calculus, it is often possible to obtain negative solutions that are not appropriate to the problem posed. The final step of looking back also entails making sure that we have actually addressed the question asked or the problem posed, rather than a different, somewhat related, question or problem.

Although these four phases appear linear, the majority of mathematical problems require various iterations of these phases. In particular, it is often the case that the first plan created to solve a problem does not work, so it has to be revised following an attempt to carry it out. Good problem solvers see these revisions as opportunities to learn more about the problem, rather than as failure at the problem solving process.

1.3.2 Modeling with Mathematics

Mathematical problem solving often relies on modeling with mathematics, particularly in the context of real-world phenomena. The Guidelines for Assessment and Instruction in Mathematical Modeling Education (GAIMME) defines mathematical modeling as "a process that uses mathematics to represent, analyze, make predictions or otherwise provide insight into real-world phenomena" [Garfunkel and Montgomery, 2019, p. 8]. This definition varies from other common connotations of "modeling" that are used in education, including in mathematics. For example in mathematics education, using manipulatives to model a mathematical idea; sketching graphs or pictures to communicate or understand concepts; and demonstrating how to solve certain types of problems are all referred to as modeling. However, as a process standard, modeling mathematics does not generally include these activities. Instead, it refers to activities of using mathematics to analyze, predict, and represent real-world data.

The GAIMME Report breaks the mathematical modeling process into the six components shown in Figure 1.3. Critically, these six components **do not** happen in a progression, but instead are iterative and sometimes run parallel to each other.

The focus on trying to better understand real-world phenomena distinguishes mathematical modeling from application problems or word problems. Actual real-world phenomena are distinctly messy and arriving at a final solution requires interpretation and assumptions. Moreover, an individual engaged in modeling has to confront the challenge of determining whether the question, mathematical model, and data are

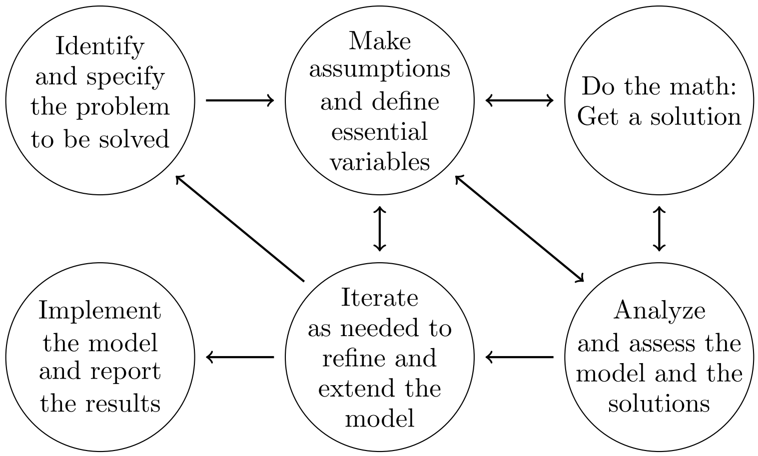


Figure 1.3: The Math Modeling Process [@GAIMME, p. 13]

well-aligned. Teachers face the additional challenge of ensuring that the situations they create for their students are appropriate for their students. For instance, an elementary school classroom might plant some seeds in the soil. The teacher would then assist the students in identifying specific questions that can be answered quantitatively about the seeds and plants, guiding them towards a set of questions and data that the children could collect, analyze, and interpret. A secondary teacher teaching about exponential functions may introduce the class to the concept of population growth and guide the students to questions that can be answered quantitatively with exponential models. In both these examples, the instructor introduces the students to a real-world situation and then guides them to questions that facilitate the students' development of the mathematical modeling they want to discuss.

Another key aspect of the modeling process involves determining the relevance of various quantities in the situation and how to describe them with variables. This cyclical process involves identifying variables, assigning them labels and then assessing how they fit in with other variables. This process allows the modeler to create an idealized version of the original problem in order to create some type of solution. It is important during the process to justify each assumption made and to clearly label each variable, along with its appropriate units. It is through this justification process that the modeler can communicate the applicability and interpretation of the results from the model to the consumer.

As the modeler defines the variables and the relationships between them, they can use the mathematical problem solving techniques to 'do the math' and come up with possible solutions to the problems posed. This section of 'doing the math' reflects the word problems of most math textbooks since the word problems rarely have extraneous information and have obvious variables defined with transparent relationships between them.

Similar to the 'looking back' phase of the mathematical problem solving process, the modeling process includes a stage during which the modeler steps back and assesses the model. During this stage he or she analyzes and tests the model and solution, examining the model created and determining the appropriateness and accuracy of the solution produced. This process often results in a revision to the original assumptions and adjustments to the model with improved variables and relationships between the variables.

Since the mathematical modeling process deals with messy real-world phenomena, an iterative process of reflection that results in the refinement and, where appropriate, extension of the the model is a key component that almost always appears in a modeling cycle. This process may run parallel to or between any of the other phases of the modeling process.

When the model gets to a state that satisfies the modeler in terms of the model justifications and appropriateness of the solutions, the implementation of the model and the reporting of the results occurs. The presentation of the results must include the justification for the model, a description of the various assumptions made to produce the model, and any limitations the model has in terms of the accuracy of the results.

This presentation must be done in a way that is comprehensible to the desired audience. Because modeling deals with real-world situations, the solution is unlikely to be clear or definitive, but will instead be approximate and estimated.

1.3.3 Communicating Mathematically

Mathematical communication broadly relates to the ways and methods of representing, justifying, and interpreting mathematics. More specifically, the practice of mathematical communication "encompasses both listening and reading (comprehension) and both speaking and writing (expression)... and may also include representation of mathematical ideas in nonlinguistic ways" [Sammons, 2018, p. 7]. In order to better understand the practice of communicating mathematically, we examine six overlapping concepts that are at the core of this practice (see Figure 1.4).

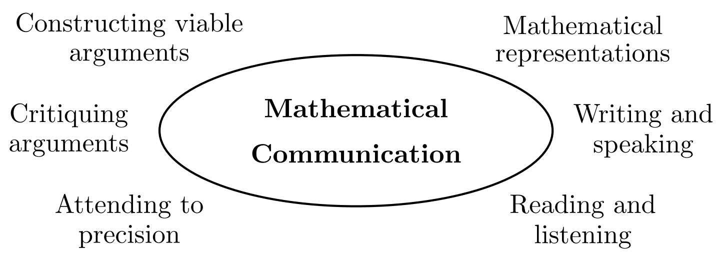


Figure 1.4: Components of Mathematical Communication

All types of communication require a language. Within a language there exists a shared meaning behind the words and symbols used when communicating with a particular language. Mathematics communication, in this sense, operates much as a language does. However, unlike a language such as English, mathematics is represented using a mixture of words that have very precise definitions and symbols that compress many different ideas and meanings into a very small amount of space (linguists call mathematics a "dense" language because of this property). For example, the sentence

$$\{f : \mathbb{C} \to \mathbb{C} | f(z) = az + b,\ a, b \in \mathbb{C}\}$$

is very concise in terms of symbols, but contains a large amount of mathematical language and concepts that the average person may not understand.

Moreover, somewhat uniquely to mathematics, the meaning of symbols varies widely depending upon the context of the communication. Consider variables, which are usually represented using any number of symbols. While certain types of symbols will cue a well-informed mathematics student that a variable has some implied meaning in terms of what type of number system to which it belongs, even in this case variables in different circumstances serve different roles. Sometimes the letter represents a fixed value (e.g., $x = 4$) and other times it represents a range of values that satisfy a particular condition (e.g., $x \leq 4$). Sometimes it indexes a set, while other times it represents a function. The differences between these uses are often subtle, making explicit training in these nuances mathematics critical.

Once the language of mathematics has been developed, the grammar is created. In mathematics, logic, argument, and proof form the core patterns of organization and presentation. Together, logic, argument, and proof boil down to the ability to construct viable arguments and critique the arguments of others using mathematical evidence.

Part of mathematical communication is learning how to consume and understand the ideas of others. In mathematics this usually occurs by reading or listening to the arguments of others. Both reading and listening to mathematics require a great deal of practice and are skills students must hone over many courses. In particular, reading mathematics requires a great deal of practice and discipline. While a novel can usually be consumed at a steady pace without much explicit effort by the reader, written mathematical ideas are often presented symbolically, with the reader left to fill in certain logical steps. Thus, the act of reading mathematics becomes one of translation of the symbolic work into a language he or she understands, followed by a effort to connect current parts of the argument to previously made points, and the absorption and evaluation of those ideas based on previous knowledge. As a result, reading mathematical texts and arguments requires a big investment in time. Sometimes it can take hours just to understand one line of a mathematical text! This slow, methodical way of reading is challenging to students of mathematics, but is a critical skill to develop as it is core to the practice of doing mathematics.

Previous paragraphs have described the difficulties and process of consuming mathematical texts. However, it is also important that students learn to produce their own mathematical communications in both oral and written forms. In communicating to others, it is essential that precise mathematical words are used, rather than slang, in order to reduce the confusion of potential readers. It is also useful to employ a variety of mathematical representations, including graphs, sketches, and diagrams to facilitate the communication of mathematical ideas.

As teachers of mathematics, we must continually emphasize the development of this practice of communicating mathematically by de-emphasizing simple answers to mathematical problems, teaching students how to read and write mathematics (not just about mathematics), and to promote classroom discourse to provide opportunities for students to experience different types of mathematical communication.

1.3.4 Understanding Mathematical Structures

While the definition of mathematics is not uniform or agreed on, most agree that mathematics is inherently about the study of underlying structures and logical reasoning. For example, the Encyclopædia Britannica -Fraser et al. [2019] defines mathematics as "the science of structure, order, and relation that has evolved from elemental practices of counting, measuring, and describing the shapes of objects." This means that an important practice in the doing of mathematics is to "look for and make use of structure" and to "look for and express regularity in repeated reasoning" [NGA-CCSSO, 2010].

The pursuit and use of structure and regularity appear throughout all of mathematics. Mathematical structure informs every part of mathematics teaching, from instruction on common mathematical procedures to searching for and making connections between various, apparently unrelated, mathematical content areas to provide new insight into a problem. As teachers, we need to point out these mathematical structures to the students and help them learn how to discover the structures and regularity on their own.

1.3.5 Exercises

1. Consider the following mathematical task:

   An electricity company charges Kelly \$0.15 per kWh (kilowatt-hour) of electricity used, plus a basic connection charge of \$8.00 per month. Find a function that helps Kelly estimate her monthly electricity bill for a given number of kWhs. Be explicit about the domain of the function you create.

   a) NCTM emphasizes multiple representation of functions using verbal, tabular, graphical, and symbolic representations. Using this word problem as the verbal representation, express the function in each of the remaining ways.

   b) How does the ability to move between different representations of functions support students' development of the mathematical communication process standard?

   c) The problem solving section discussed the fact that a mathematical task is only considered a problem if the person working on the task does not know in advance a solution method that will produce a correct answer. Would you say that for part (a) you were engaged in mathematical problem solving as described in the standards? Why or why not?

   d) Under what conditions would the task be a problem for students you were teaching?

e) Based on the description of mathematical modeling, would you say that finding the function for the task in part (a) represents mathematical modeling? Why or why not?

f) A second part of the task asks students to estimate Kelly's electrical bill for the year, given that her monthly kWh usage ranges between 202 and 254 kWh.

g) The textbook you got the problem from lists the answer to the extension as $506.40. What did the author do and what assumptions did she make in order to arrive at that answer? Do you agree with her process and assumptions? Why or why not?

h) In arriving at your answer for part (a), would you say that you were engaged in mathematical problem solving? Why or why not?

i) If you were using this task as a modeling activity for your students, what criteria would you use to evaluate whether their answer was reasonable?

2. There are five NCTM Process Standards and eight Common Core Standards for Mathematical Practice. While there is significant overlap between the two sets of standards, they are not the same. Read each set of standards. When there is overlap between the two sets of standards, create a map between them. Also identify the ways in which the two sets of standards differ. These regions are not at the heading level. You'll have to actually dig into the blurbs about each standard in their original documents.

3. A basic theorem that students use from an early age is that the sum of two even integers is even. A typical proof of this theorem might look something like this:

   Let a and b be even integers such that $a = 2m$ and $b = 2n$ where m and n are integers. Then $a + b = 2m + 2n = 2(m + n)$. Thus, the sum of two even integers is even.

   a) The section on mathematical communication emphasizes the reliance of mathematical communication on precise definitions and symbols that compress complex ideas into short phrases. In examining this proof, identify the places where the problem statement or proof rely on a definition or compressed expression.

   b) Identify the places where the proof writer has left the reader to fill in logical steps or rationales.

   c) Rewrite the proof of the theorem without the use of any symbols, but with the same degree of precision and generalizability.

   d) In comparing the symbolic proof to your verbal proof, do you think it is easier to understand (i.e., more like a novel?) or do you think it would still be difficult to read if all mathematics were presented without symbols? Why or why not?

   e) If you reflect on the six parts of mathematical communication, which parts do you think you are best at? Which parts do you struggle the most with? What is one goal you have for yourself for mathematical communication that you want to develop during this course?

1.4 Mathematics Content Standards in the U.S.

The current standards for mathematics and assessment in the United States are derived from a variety of sources and past initiatives. David Klein -Klein [2003] gives a good summary of the major events and influential individuals during the 20$^{\text{th}}$ century. The current set of standards are strongly connected to a pair of reports from the 1980's: one by NCTM -NCTM [1980], *An Agenda for Action*, and one by the U.S. National Commission on Excellence in Education -National Commission on Excellence in Education [1983], *A Nation at Risk*. Each provided a different view about what should occur to improve mathematics education in the United States. While these two documents were used as the basis for the 'math wars' of the late 20$^{\text{th}}$ century, the biggest difference between them is that the NCTM document focused primarily on standards for teaching methodologies, while *A Nation at Risk* focused on standards of content knowledge.

The NCTM -NCTM [1980] recommended "that

1. problem solving be the focus of school mathematics in the 1980s;
2. basic skills in mathematics be defined to encompass more than computational facility;
3. mathematics programs take full advantage of the power of calculators and computers at all grade levels;
4. stringent standards of both effectiveness and efficiency be applied to the teaching of mathematics;
5. the success of mathematics programs and student learning be evaluated by a wider range of measures than conventional testing;
6. more mathematics study be required for all students and a flexible curriculum with a greater range of options be designed to accommodate the diverse needs of the student population;
7. mathematics teachers demand of themselves and their colleagues a high level of professionalism;
8. public support for mathematics instruction be raised to a level commensurate with the importance of mathematical understanding to individuals and society" (p. 1).

The National Commission on Excellence in Education -National Commission on Excellence in Education [1983] recommended that

- "The teaching of *mathematics* in high school should equip graduates to: (a) understand geometric and algebraic concepts; (b) understand elementary probability and statistics; (c) apply mathematics in everyday situations; and (d) estimate, approximate, measure, and test the accuracy of their calculations. In addition to the traditional sequence of studies available for college-bound students, new, equally demanding mathematics curricula need to be developed for those who do not plan to continue their formal education immediately" (p. 25).
- "Standardized tests of achievement (not to be confused with aptitude tests) should be administered at major transition points from one level of schooling to another and particularly from high school to college or work. The purposes of these tests would be to: (a) certify the student's credentials; (b) identify the need for remedial intervention; and (c) identify the opportunity for advanced or accelerated work. The tests should be administered as part of a nationwide (but not Federal) system of State and local standardized tests. This system should include other diagnostic procedures that assist teachers and students to evaluate student progress" (p. 28).
- "Persons preparing to teach should be required to meet high educational standards, to demonstrate an aptitude for teaching, and to demonstrate competence in an academic discipline. Colleges and universities offering teacher preparation programs should be judged by how well their graduates meet these criteria" (p. 30).
- "Substantial nonschool personnel resources should be employed to help solve the immediate problem of the shortage of mathematics and science teachers. Qualified individuals including recent graduates with mathematics and science degrees, graduate students, and industrial and retired scientists could, with appropriate preparation, immediately begin teaching in these fields. A number of our leading science centers have the capacity to begin educating and retraining teachers immediately. Other areas of critical teacher need, such as English, must also be/addressed" (p. 31).

NCTM followed it's 1980 report with the publication of the *Curriculum and Evaluation Standards for School Mathematics* [NCTM, 1989]. This document focused on the standards for teaching methodologies and mathematical practices. Content standards were sketched out, with overviews of the recommended content knowledge for students in 4-year grade bands. NCTM expanded this work with additional texts focusing on teaching standards and assessment: *Professional Teaching Standards* [1991] and *Assessment Standards* [1995]. In 2000, NCTM released an updated version the 1989 standards with the *Principles and Standards for School Mathematics* [NCTM, 2000]. During this same time, many states developed more specific content standards reflecting the guidelines recommended by the 1989 document [Raimi and Braden, 1998]. Although the NCTM document did not make content recomendations by specific grade levels, many of the state standards still resided at the grade-band level [Reys and Lappan, 2007, p. 677].

1.4.1 Common Core State Standards-Math

In 2002, Congress passed the *No Child Left Behind Legislation.* This bill required that states determine measurable content standards for each grade level and develop assessments based on these standards to be given to all students at specific grade levels (generally fourth, eighth, and twelfth grade) in order to receive federal school funding. Since each state developed their own standards, the level and specificity of these standards varied greatly between States [Reys and Lappan, 2007].

It was in this environment that the National Governors Association Center for Best Practices and the Council of Chief State School Officers launched an effort in 2009 to develop "a common core of internationally benchmarked standards in math and language arts for grades K-12" [NGA-CCSSO, 2010]. These standards emphasize content knowledge, while also encouraging some of the teaching methodologies proposed by the NCTM documents. However, these standards to do instruct teachers how to teach or what curriculum to use. They instead focus on what students should know and be able to do at each grade level.

In order to help teachers better understand how the topics in this book relate to the actual teaching and learning of their students, we include related content standards from the Common Core State Standards for Mathematics throughout the text. While not all states have adopted the Common Core State Standards for their school systems, these standards are a good representation of what is expected from students at different stages in their mathematical education.

Related Content Standards

- (8.F.1) Understand that a function is a rule that assigns to each input exactly one output. The graph of a function is the set of ordered pairs consisting of an input and the corresponding output.
- (HSF.IF.7) Graph functions expressed symbolically and show key features of the graph, by hand in simple cases and using technology for more complicated cases.

Standards that align with the content of each section are provided in blocks with the title "Related Content Standards." One such box is given above as an example. The sequence of numbers and letters proceeding each standard is its standardized abbreviation. The first part of the abbreviation identifies the grade band. Since the Common Core State Standards for high school do not have a recommended grade level for each of the standards, they are denoted by "HS" and the general area of mathematics. The second part of the abbreviation refers to the general domain of mathematics, and the third part refers to the location of the standard in that grade band and domain.

1.4.2 NCTM-CAEP Standards

Another requirement of the No Child Left Behind legislation was that states had to create standards for teachers to receive certification. These standards for certification needed to be measured with a blend of standardized assessments, portfolios, and coursework. As a result, certification in secondary mathematics in most states requires a major in mathematics (or its equivalent), the teaching of and about standards during certification coursework, and achievement of certain scores on a content specific standardized test (such as the PRAXIS). In addition, the NCTM partnered with the Council for the Accreditation of Educator Preparation (CAEP) to develop standards for what beginning secondary mathematics teachers should know and be able to do in the teaching of mathematics [NCTM, 2020].

The NCTM-CAEP content standards are built upon the standards developed by the Conference Board of Mathematical Sciences in -Conference Board of the Mathematical Sciences [2001] and -Conference Board of the Mathematical Sciences [2012]. The content of this book builds upon the contents of the first standard of "Knowing and Understanding Mathematics". In particular:

> Candidates demonstrate and apply, with the incorporation of mathematics technology, conceptual understanding, procedural fluency, and factual knowledge of major mathematical domains:

> Number; Algebra and Functions; Statistics and Probability; Geometry, Trigonometry, and Measurement; Calculus; and Discrete Mathematics.

We list the details of the NCTM-CAEP content standards below and discuss how this book connects to each of these standards.

> **Essential Concepts in Number.** Candidates demonstrate and apply conceptual understanding, procedural fluency, and factual knowledge of number including flexibly applying procedures, using real and rational numbers in contexts, developing solution strategies, and evaluating the correctness of conclusions. Major mathematical concepts in Number include number theory; ratio, rate, and proportion; and structure, relationships, operations, and representations.

Number systems are primarily covered in Chapters 4, 6, and 7. We introduce initial concepts of the number systems in Chapter 4, focusing on operations and the basic properties of common number systems, including whole numbers, integers, rational, real, and complex numbers. Chapters 6 and 7 expand these ideas through the added structure of rings and fields.

> **Essential Concepts in Algebra and Functions.** Candidates demonstrate and apply understandings of major mathematics concepts, procedures, knowledge, and applications of algebra and functions including how mathematics can be used systematically to represent patterns and relationships including proportional reasoning, to analyze change, and to model everyday events and problems of life and society. Essential Concepts in Algebra and Functions include algebra that connects mathematical structure to symbolic, graphical, and tabular descriptions; connecting algebra to functions; and developing families of functions as a fundamental concept of mathematics. Additional Concepts should include algebra from a more theoretical approach including relationship between structures (e.g., groups, rings, and fields) as well as formal structures for number systems and numerical and symbolic calculations.

Many of these concepts are spread throughout the text, as algebra and functions are fundamental to the secondary curriculum. That said, Chapters 5, 6, 7, and 8 specifically focus on topics of particular importance in understanding algebra and functions. In particular Chapter 8 combines the themes of the earlier chapters together using real-valued functions.

> **Essential Concepts in Calculus.** Candidates demonstrate and apply understandings of major mathematics concepts, procedures, knowledge, and applications of calculus including the mathematical study of the calculation of instantaneous rates of change and the summation of infinitely many small factors to determine some whole. Essential Concepts in Calculus include limits; continuity; the Fundamental Theorem of Calculus; and the meaning and techniques of differentiation and integration.

The essential concepts in calculus are currently excluded from this text with the understanding that most pre-service teachers are already required to take at least two semesters worth of calculus courses that focus on this content standard.

> **Essential Concepts in Statistics and Probability.** Candidates demonstrate and apply understandings of statistical thinking and the major concepts, procedures, knowledge, and applications of statistics and probability including how statistical problem solving and decision making depend on understanding, explaining, and quantifying the variability in a set of data to make decisions. They understand the role of randomization and chance in determining the probability of events. Essential Concepts in Statistics and Probability include quantitative literacy; visualizing and summarizing data; statistical inference; probability; and applied problems.

The essential concepts of statistics and probability are the focus of Part IV of the text, Data Analysis. In that section we focus on statistics and a study of variability, looking at different ways to measure, communicate, and understand variability in the context of statistical problems. We also discuss general principles of data analysis studies including data collection, hypothesis testing, and methods of reporting results.

> **Essential Concepts in Geometry, Trigonometry, and Measurement.** Candidates demonstrate and apply understandings of major mathematics concepts, procedures, knowledge, and applications of geometry including using visual representations for numerical functions and relations, data and statistics, and networks, to provide a lens for solving problems in the physical world. Essential Concepts in Geometry, Trigonometry, and Measurement include transformations; geometric arguments; reasoning and proof; applied problems; and non-Euclidean geometries.

Part III on Geometry looks at the subject of Geometry from the perspectives of constructional, transformational, analytic, and algebraic. Each perspective helps us to better understand the essential concepts in geometry, trigonometry, and measurement, along with the interactions between geometry, algebra, functions, and number systems.

1.4.3 Exercises

1. Reflect on your K-12 mathematics education. Based on the brief history described, what documents were being used to guide your curriculum?

2. Those who study mathematics curriculum describe the changing focus of what is written in school standards for mathematics as a pendulum. Over time, the pendulum swings between a focus on procedural fluency such as *A Nation at Risk* -National Commission on Excellence in Education [1983] describes and the more process oriented standards as outlined by NCTM (1989).

   a) How does the Common Core State Standards attempt to unify the two extremes of the pendulum swing?
   b) How does treating the process standards as separate from the content standards provide an opportunity to continue to let the pendulum swing?
   c) Regardless of which documents were currently in vogue while you were in school, do you think your mathematics education was more procedurally focused or more process standard focused?
   d) Is how you were taught what you hope your own teaching will be like? Why or why not?
   e) K-12 mathematics students often think that mathematics is a set of unrelated procedures that they have to memorize how to do. In contrast, people who do mathematics for a living think mathematics is a conceptual and logical system where everything fits together. As you currently understand them, do you think the Common Core State Standards could support students in moving from thinking about mathematics as a set of unrelated procedures to a more conceptual system? Why or why not?

Chapter 2

Set Theory

In the mid-1800s the field of mathematics went through a major shift that ended up changing the very definition of mathematics. In 1847, George Boole wrote in the introduction to *The Mathematical Analysis of Logic* that up to that point, "the abstractions of the modern Analysis, not less than the ostensive diagrams of the ancient Geometry, have encouraged the notion, that Mathematics are essentially, as well as actually, the Science of Magnitude" [Boole, 1847]. Instead, Boole proposed a new definition, suggesting

> We might justly assign it as the definitive character of a true Calculus, that it is a method resting upon the employment of Symbols, whose laws of combination are known and general, and whose results admit of a consistent interpretation... It is upon the foundation of this general principle, that I purpose to establish the Calculus of Logic, and that I claim for it a place among the acknowledged forms of Mathematical Analysis, regardless that in its object and in its instruments it must at present stand alone.

Figure 2.1: George Boole

At the time that Boole wrote the above passage, mathematics as a field shifted from the study of quantities to the study of abstract structures based on logic and set theory. This change in the definition of mathematics freed up those who studied it to move beyond structures tied to physical interpretations and applications and move to a more abstract field. The abstractification of mathematics was a powerful moment, resulting in the development of fields such as quantum mechanics, relativity, cryptography, statistics.

In this chapter we will go through some of the basics of set theory needed to understand some of the later material and to develop a common vocabulary and set of notations. We do not offer a deep treatment of set theory as there are many textbooks, specifically in the area of Discrete Math, with a more detailed coverage of it.

2.1 Sets and Subsets

The foundation of modern mathematics is the theory of sets. Informally, sets can be thought of as collections of objects. While **set theory** is not specifically outlined in the content standards of most states, it is mentioned in the Common Core standards. The first reference to set theory occurs in the Introduction to Kindergarten, stating:

> Students use numbers, including written numerals, to represent quantities and to solve quantitative problems, such as counting objects in a set; counting out a given number of objects; comparing sets or numerals; and modeling simple joining and separating situations with sets of objects, or eventually with equations such as $5 + 2 = 7$ and $7 - 2 = 5$. (Kindergarten students should see addition and subtraction equations, and student writing of equations in kindergarten is encouraged, but it is not required.) Students choose, combine, and apply effective strategies for answering quantitative questions, including quickly recognizing the cardinalities of small sets of objects, counting and producing sets of given sizes, counting the number of objects in combined sets, or counting the number of objects that remain in a set after some are taken away. [NGA-CCSSO, 2010]

Set theory also makes an appearance in the high school standards related to counting and probability, as knowledge of sets is critical to understanding basic formulas in probability.

Related Content Standards

- (HSS.CP.1) Describe events as subsets of a sample space (the set of outcomes) using characteristics (or categories) of the outcomes, or as unions, intersections, or complements of other events ("or," "and," "not").

We start our exploration of set theory by developing some basic notation and definitions. Cantor [1891] defined sets in the following way.

> Unter einer 'Menge' verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objecten m unsrer Anschauung oder unseres Denkens (welche die 'Elemente' von M genannt werden) zu einem Ganzen. (p. 481)
>
> By a 'Set' we mean each collection M of certain well-differentiated objects m of our perception or our thinking (which are called the 'elements' of M) as a whole. (English translation)

This definition proved to not be precise enough to avoid certain paradoxes.

Example 2.1 (Russell's Paradox)**.** Let $R = \{x | x \notin x\}$, the set that contains all sets that are not members of themselves. Is R an element of R? If it is, then it is a set that is an element of itself and so could not be an element of R, by its definition, thus leading to a contradiction. On the other hand, if one assumes that R is not an element of itself, then it satisfies the conditions of being an element of R, also leading to a contradiction. Thus defining the paradox.

Even though Cantor's definition of a set leads to such paradoxes, we will use it as our working definition, with the detailed definition of set being a primitive notion in Zermelo-Fraenkel set theory. The Zermelo-Fraenkel axioms, combined with the axiom of choice, create the ZFC axioms upon which mathematics is built. Due to the level of abstraction involved in the axioms, we will not include a systematic coverage of the axioms in this text, but will reference them as needed.

Definition 2.1. Set Definitions:

- We will define a set by the collection of elements which belong to the set.
- If A is a set and a is an object that belongs to A, we say that a is an **element** of A, denoted as $a \in A$.
- If an object a is not an element of a set A, denoted as $a \notin A$.
- Let A and B be sets. We say that $A = B$ if and only if every element of A is an element of B and every element of B is an element of A. (In the ZFC axioms, this is referred to as the axiom of extensionality.)

While the elements of a set are often written in a specific order, i.e. $\{1, 2, 3\}$, the members of a set have no particular order and the same set could be written as $\{3, 1, 2, 1\}$, with the repetition being irrelevant since the set is defined by its elements.

We also need to note that a and $\{a\}$ are two different mathematical objects (one is the element a and the other is the set that contains the element a). So $A = \{a, \{a\}\}$ defines a set with two distinct elements: a and $\{a\}$. Likewise, if $B = \{1, 2, 3, \{4\}, 5\}$, then $4 \notin B$ but $\{4\} \in B$. So the symbol 4 is not an element of B, but the symbol $\{4\}$ (the set that contains the element 4) does belong to B. These nuances mean that we have to be very careful with our notation and how we read the mathematical symbols.

Another way to describe sets is using *set-builder notation*. In this notation, we describe our new set using larger sets and a set of restrictions on (or description of) which objects in the larger set we are choosing to include. For instance, we can define C to be the set of all real numbers (denoted $\mathbb{R}$) greater than or equal to 3. In this case, the larger set is the set of real numbers and the condition that the numbers are greater than or equal to three is the restriction or description. Of course we do not want to have to keep writing so much down, so we create a short-hand way of defining this set:

$$C = \{x \in \mathbb{R} | x \geq 3\}$$

where the $|$ separates the description of the larger set and the description of the restrictions, and is usually read as "such that" when read aloud. This notation is read "C is the set of all of the elements x in the real numbers such that x is greater than or equal to 3." This set-builder notation is particularly useful when describing sets with more than just a few elements.

When sets are subsets of the real numbers, we can also describe them using *interval notation*. For the set C defined above, we can also write

$$C = [3, \infty)$$

where the closed bracket, [, denotes that the endpoint is contained in the set, while an open bracket, (, denotes that the endpoint is not contained in the set. For real numbers a and b, with $a < b$, we have the following options as intervals from a to b:

$$(a, b) \quad (a, b] \quad [a, b) \quad [a, b].$$

Once we have created notation for sets with a few elements and a large number of elements, we can describe how to denote a set without any elements.

Definition 2.2. The set that does not contain any elements is called the **empty set** and is denoted by $\emptyset$. Any set that contains at least one element is then called **non-empty**.

In addition to determining how to describe sets, we must have a way to determine if two sets are the same or distinct.

A direct consequence of this definition is that the number of times an element is listed and the order of the elements is irrelevant for equality of sets. So $\{a, b, c, d\} = \{b, a, c, c, d, b\}$ and $\{1, 2\} \neq \{1, 2, 3\}$.

Definition 2.3. Let A and B be sets. We say that A is a subset of B, denoted $A \subseteq B$, if every element of A is also an element of B. If A is not a subset of B, we sometimes denote this by $A \nsubseteq B$. If A is a subset of B and there are elements of B that are not contained in A, then we say that A is a **proper subset** of B and denote it by $A \subset B$.

Note that these symbols of $\subseteq$ and $\subset$ are similar to $\leq$ and $<$. While there are some overlaps in their meanings, one should be careful to recognize their similarities and differences.

For example:

$$\{6, 7\} \subseteq \{5, 6, 7, 8\}, \quad \text{or} \quad \{6, 7\} \subset \{5, 6, 7, 8\}$$

$$\{x \in \mathbb{R} | x > 5\} \subseteq \{x \in \mathbb{R} | x \geq 2\}, \text{ and}$$

$$\{a, b, c\} \subseteq \{a, b, c\}.$$

Notice that the first two examples are proper subsets. When we have proper subsets, we often use the $\subseteq$ notation rather than $\subset$ when the property of not being the same set is not essential.

It is important to distinguish between the phrases 'element of' and 'contained in' when discussing sets. If $A = \{a, b, c\}$, then we say that a is an element of A, while $\{a\}$ is contained in A

Since the empty set has no elements, it is by default a subset of every set. Similarly, every set is a subset of itself.

Example 2.2. If $A = \{a, b, c\}$ we can describe all of the subsets of A,

$$\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}.$$

This set of subsets of A is called the **power set** of A and denoted by $\mathcal{P}(A)$.

We often need to prove that two sets are equal to one another, and we do not have the elements of the set listed out. In these situations the following theorem often proves useful.

Theorem 2.1. *Let A and B be sets. $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.*

Proof. Since the statement of the theorem is an 'if and only if' statement (sometimes denoted 'iff'), there are actually two components that need to be proven. We need to first prove that if $A = B$ then $A \subseteq B$ and $B \subseteq A$. The second statement that needs proving is the converse: having $A \subseteq B$ and $B \subseteq A$ implies that $A = B$. We will complete both arguments using the corresponding definitions.

Assume that $A = B$. Then the definition of set equality states that every element of A is an element of B. This means that A meets the definition of a subset of B. That is: $A \subseteq B$. In addition, the assumption that $A = B$ gives us that every element of B is an element of A. As a result, $B \subseteq A$. Thus

$$A = B \Rightarrow A \subseteq B \text{ and } B \subseteq A.$$

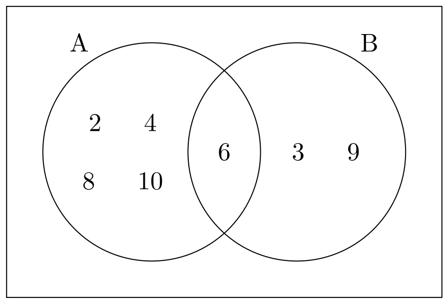
To prove the converse statement, we assume that $A \subseteq B$ and $B \subseteq A$. Thus, we have that every element of A is also an element of B and every element of B is also an element of A. This is the definition of set equality and so we have that $A = B$.

We have therefore proven both implications, and thus the theorem is proven. $\square$

2.1.1 Venn Diagrams

In order to better understand relationships between subsets of a larger set, it is sometimes helpful to represent the relationships with a **Venn diagram**. In these diagrams, we define a 'universe' in which the sets reside, denoted by a rectangle. We use circle-like figures to represent sets inside that universe, with everything inside a circle being inside of the corresponding set and everything outside of a circle being outside of the corresponding set, but inside the related universe.

Example 2.3. Let $A = \{2, 4, 6, 8, 10\}$ and $B = \{3, 6, 9\}$ inside of the universe of natural numbers, $\mathbb{N}$. Then these sets could be represented by a Venn diagram such as the one below.



As previously discussed, sets do not have to just be composed of numbers. Moreover, sets can have many different relationships between them, as seen in Figure 2.2. If a set is a subset of a second set, then they are visualized as one region inside of the other region. If there are elements shared between two sets, but are not known to have a subset relationship, then they are visualized as overlapping regions. If no elements of the two sets are the same, then they are represented as two non-overlapping regions. In the next section we define vocabulary to describe these different types of relationships and how they are combined in different ways.

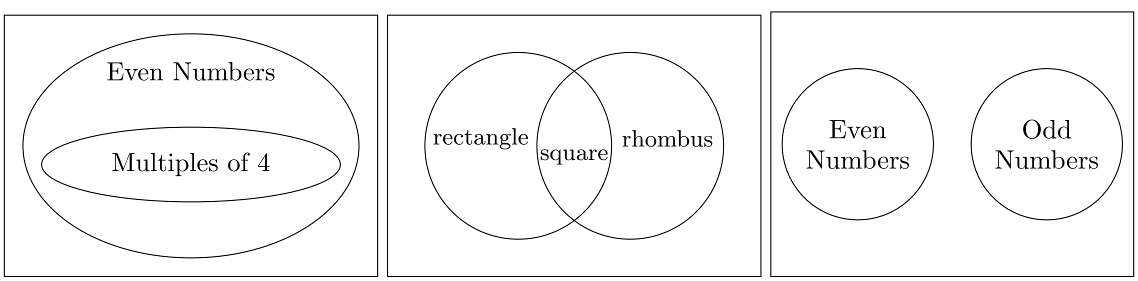


Figure 2.2: Sample Venn diagrams

2.1.2 List of Sets of Numbers

In order to help with terminology we will provide some notation for some of the basic sets used in the text.

We define the natural numbers as

$$\mathbb{N} = \{0, 1, 2, 3, 4, 5, \ldots\}.$$

This is distinct from some definitions of the natural numbers that do not include the number 0. When a text defines the natural numbers without the 0, it also defines the whole numbers to be our definition of the natural numbers. We are including 0 due to the method in which we define the natural numbers in Chapter 4.

We label the integers as

$$\mathbb{Z} = \{\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots\},$$

which we will define in detail in Section 4.1.

We label the set of rational numbers as $\mathbb{Q}$ and the real numbers as $\mathbb{R}$, the detailed definitions of which are in Sections 4.4 and 4.6, respectively.

The complex numbers, $\mathbb{C}$, are defined and studied in Section 4.7.

For the integers, rationals and reals, we define $\mathbb{Z}^+$, $\mathbb{Q}^+$, and $\mathbb{R}^+$ to be the positive elements of the set, those that are greater than zero.

We visualize the nested nature of these number systems in the Venn diagram in Figure 2.3.

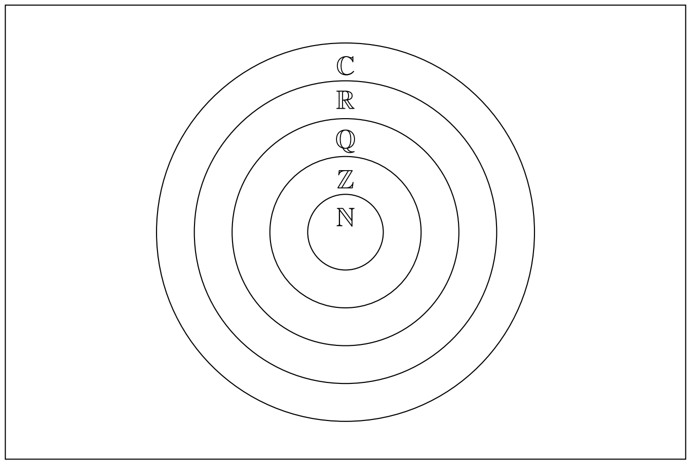


Figure 2.3: Sample Venn diagrams

2.1.3 Exercises

1. Answer the following as true or false. If false, explain why the statement is not true.

    a. $\emptyset \subseteq \{f, u, n, t, i, m, e, s\}$
    b. $\{a, b\} \in \{a, b, c\}$
    c. $\{0\} \ = \emptyset$
    d. $\{f, u, n, f, u, n\} = \{f, u, n\}$
    e. $\{0, 0\} \subseteq \{0, 0, 1, 1, 2, 2\}$

2. Let $A = \{1, 2, 3\}$, $B = \{a, b, c\}$, and $C = \{1, a, 2, b, 3, c\}$. Answer the following as true or false. If false, explain why the statement is not true.

    a. $A \subseteq A$
    b. $A \supseteq C$
    c. $A = B$
    d. $A \nsubseteq B$
    e. $B \subset C$
    f. $C \subset B$

3. Recall that for set A, $\mathcal{P}(A)$ is the power set of A.

    a. Let $A = \{a, b\}$. Write the set of $\mathcal{P}(A)$ by listing its elements.

b. Let A be a set with n elements in it.
   i. How many elements are in $\mathcal{P}(A)$ if $n = 3$? (For example, $A = \{a, b, c\}$)
   ii. How many elements are in $\mathcal{P}(A)$ if $n = 4$? (For example, $A = \{a, b, c, d\}$)
   iii. How many elements are in $\mathcal{P}(A)$ if n is an unknown natural number?

4. Some textbooks describe a set as "a well-defined collection of objects" which means that the inclusion criteria that helps you decide what should be in the set is clearly specified. In particular, the universe that the set is contained in needs to be clear. Classify each of the following sets as well-defined or not. If you identify a set as not well-defined, give two possible well-defined sets that would satisfy the original description.

   a. $\{x|\, x > 0\}$
   b. The set of students at The University of Awesome who are currently enrolled in a class that has a 100-level designation.
   c. $\{x|\, x$ is a letter in my first name$\}$
   d. The set of my friends.
   e. $A_n = \{x \in \mathbb{Z}|\, n \leq x \leq n + 3\}$

5. Let $A = \{1, 2, 3, 4\}$, $B = \{7, 8, 9\}$, and $C = \{3, 4, 5, 6, 7, 8, 9\}$. Using the most appropriate template of the Venn diagrams shown in Figure 2.2, fill in the regions with the elements from:

   a. Sets A and B.
   b. Sets A and C.
   c. Sets B and C.
   d. Sets A, B, and C (you will have to make a new Venn diagram template).

6. What are possibilities for the universal sets that are implicitly defined in the examples in Figure 2.2.

2.2 Algebra of Sets

Now that we understand the basic definitions involving sets, we examine set operations. These will allow us to create new sets from given sets.

Definition 2.4. Let A and B be sets. We define the union of A and B, denoted $A \cup B$, to be the set of elements that are either in A or B, or possibly both. We define the intersection of A and B, $A \cap B$, to be the set of elements that are in both A and B. We say that A and B are **disjoint** (or **mutually exclusive**) if $A \cap B = \emptyset$.

Note that in the English language that the word 'or' is often understood in an exclusionary way. If someone asks if you would like red grapes or green grapes, there is an implication that you have to choose one or the other, but not both. In mathematics, the word 'or' is an inclusive word with the implication that you can be in both sets.

The shaded regions in Figure 2.4 illustrate each set relationship in terms of general sets A and B. We can also write the union and intersection of sets in terms of set-builder notation:

$$A \cup B = \{x|x \in A \text{ or } x \in B\} \quad \text{and} \quad A \cap B = \{x|x \in A \text{ and } x \in B\}.$$

Let's look at a couple of examples to better understand these unions and intersections.

Example 2.4. Let $A = \{a, b, c, d, e\}$ and $B = \{d, e, f, g\}$. Then

$$A \cup B = \{a, b, c, d, e, f, g\}$$

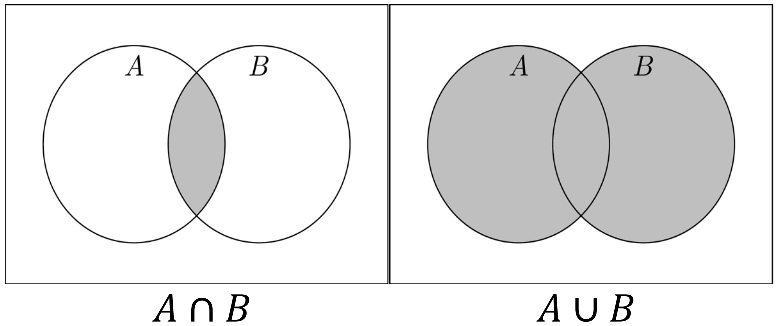


Figure 2.4: The union and intersection of two sets

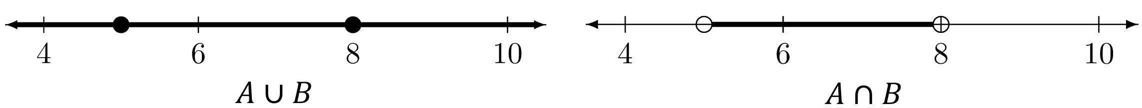
Example 2.5. Let $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $B = \{2, 4, 6, 8\}$. Then

$$A \cup B = A \quad \text{and} \quad A \cap B = B.$$

Example 2.6. Let $A = \{a, b, c\}$ and $B = \{1, 2, 3\}$. Then

$$A \cup B = \{a, b, c, 1, 2, 3\} \quad \text{and} \quad A \cap B = \emptyset.$$

Example 2.7. Let $A = \{x \in \mathbb{R} \mid x > 5\}$ and $B = \{x \in \mathbb{R} \mid x < 8\}$, also denoted as $(5, \infty)$ and $(-\infty, 8)$, respectively. Then $A \cup B$ would be all real numbers, since any real number is either less than 8 or greater than 5. And $A \cap B$ would be the real numbers between 5 and 8. These can also be represented on number lines, where the shaded lines and filled dots are included in the set, with open dots and unshaded lines not being included in the set.



The following theorem identifies properties of the union as an operation on sets that follow from the definition.

Theorem 2.2. *Let A, B, and C be sets. Then we have the following:*

1. $A \cup \emptyset = A$
2. $A \cup A = A$
3. $A \cup B = B \cup A$
4. $A \cup (B \cup C) = (A \cup B) \cup C$
5. $A \subseteq A \cup B$
6. *If $A \subseteq B$, then $A \cup B = B$*

Proof. The first two statements follow directly from the definition of the union. The empty set has no elements, making its union with A equal to A. Because A and A have the same elements, the union must also be A.

The third statement suggests that the set operation "union" is commutative, in that the order of the operation does not matter. The proof follows from properties of symbolic logic in relation to the use of the "or" statement in the definition of the union.

The fourth statement in the theorem is used to expand the definition of union, which is only defined for a pair of sets, to more than two sets. A further consequence of statement four is that the method of pairing sets the under the operation of union is irrelevant. We show this by proving the two statements $A \cup (B \cup C) \subseteq (A \cup B) \cup C$ and $(A \cup B) \cup C \subseteq A \cup (B \cup C)$ (see Theorem 2.1). In order to show the first containment, we let $x \in A \cup (B \cup C)$ be a generic element. Then by the definition of the union of sets, $x \in A$ or $x \in (B \cup C)$. We will then break this statement into two cases.

Case 1. If $x \in A$, then by the definition of unions $x \in (A \cup B)$. Then using the definitions of unions again, $x \in ((A \cup B) \cup C)$.

Case 2. If $x \in (B \cup C)$, then $x \in B$ or $x \in C$. If $x \in B$, then $x \in (A \cup B)$ and also $x \in ((A \cup B) \cup C)$. If $x \in C$, then $x \in ((A \cup B) \cup C)$.

So in either case, $x \in A \cup (B \cup C)$ implies that $x \in (A \cup B) \cup C$. So $A \cup (B \cup C) \subseteq (A \cup B) \cup C$. The proof of the reverse containment is nearly identical.

We leave the last two statements as exercises. □

Similar to the union, the following theorem demonstrates how the intersection of sets works as an operation on sets. The proofs for each statement are similar to that of the unions so we will leave the proof of this theorem as an exercise.

Theorem 2.3. *Let A, B, and C be sets. Then we have the following*

1. $A \cap \emptyset = \emptyset$
2. $A \cap A = A$
3. $A \cap B = B \cap A$
4. $A \cap (B \cap C) = (A \cap B) \cap C$
5. *If $A \subseteq B$, then $A \cap B = A$*

Now that we know how unions and intersections behave by themselves, we examine how they interact with each other.

Theorem 2.4. *Let A, B, and C be sets. Then*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{and} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

These relationships are diagrammed in Figure 2.5.

Proof. As each of these are proofs of equality of sets, we will need to complete the proofs showing that each set is contained in the other (see Theorem 2.1). We prove Part 1 and leave Part 2 as an exercise.

Let $x \in A \cap (B \cup C)$. Then $x \in A$ and $x \in (B \cup C)$. This yields two cases: $x \in A$ and $x \in B$, or $x \in A$ and $x \in C$. This is equivalent to $x \in (A \cap B) \cup (A \cap C)$. Thus $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

If $x \in (A \cap B) \cup (A \cap C)$, then we have to examine two cases: $x \in (A \cap B)$ or $x \in (A \cap C)$.

Case 1. If $x \in (A \cap B)$, then $x \in A$ and $x \in B$. Since $x \in B$, we know that $x \in (B \cup C)$. Thus $x \in A \cap (B \cup C)$.

Case 2. If $x \in (A \cap C)$, then $x \in A$ and $x \in C$. Since $x \in C$, we know that $x \in (B \cup C)$. Thus $x \in A \cap (B \cup C)$.

These results imply that $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. Thus, the Part 1 of the theorem is proven. □

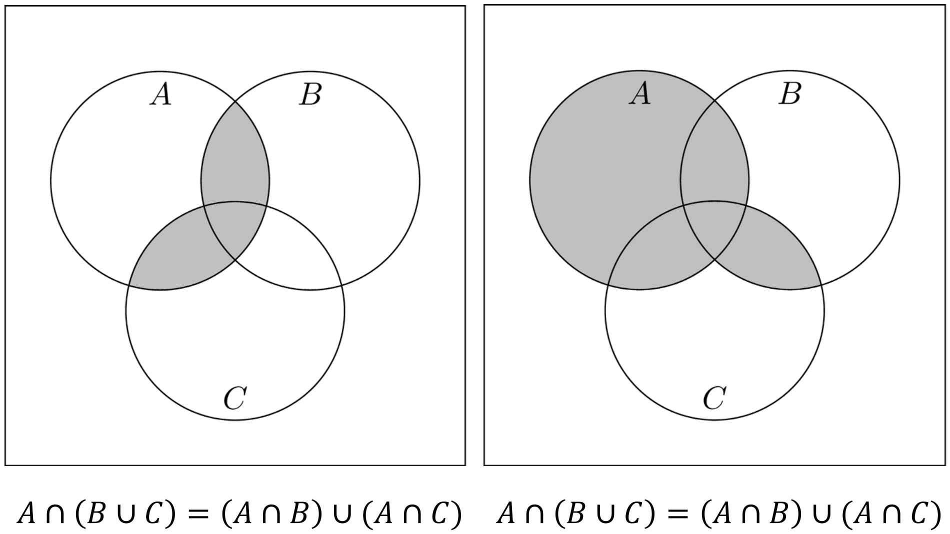


Figure 2.5: Venn Diagrams for Set Distribuions

2.2.1 Set Complements

When working on a problem, we usually describe several sets, with an underlying assumption that the sets referenced contain elements from some common, larger set. We call a set that contains all of the elements considered for a particular situation a **universal set**.

Definition 2.5. For every set A that is a subset of the universal set U, we define the **complement** of A to be the set of elements in the universal set that are not in A, denoted

$$A^c = \{x \in U | x \notin A\}.$$

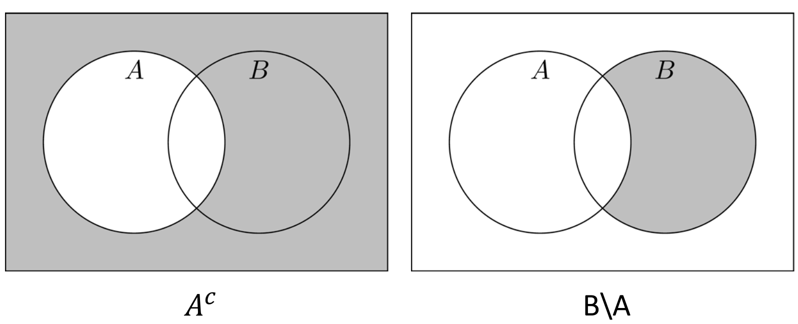


Figure 2.6: Set Complement and Set Difference

For a given set A in universal set U, the complement of a set identifies everything that is in the universal set except for things in set A. This is often useful; however, there are times when it is important to consider the elements that are in one set, but not in another, without reference to the universal set.

Definition 2.6. Let A and B be sets. The complement of A with respect to B, also called the **set difference** of B and A, is defined as

$$B \setminus A = B \cap A^c = \{x \in B | x \notin A\}.$$

Note that this can be read of as "B, not including A" or as "B minus A."

Let's revisit our previous examples to understand this idea of a set difference.

Example 2.8. Let $A = \{a, b, c, d, e\}$ and $B = \{d, e, f, g\}$. Then

$$A \setminus B = \{a, b, c\} \quad \text{and} \quad B \setminus A = \{f, g\}.$$

Example 2.9. Let $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and $B = \{2, 4, 6, 8\}$. Then

$$A \setminus B = \{1, 3, 5, 7, 9\} \quad \text{and} \quad B \setminus A = \emptyset.$$

Example 2.10. Let $A = \{a, b, c\}$ and $B = \{1, 2, 3\}$. Then

$$A \setminus B = A \quad \text{and} \quad B \setminus A = B.$$

2.2.2 De Morgan's Laws

In the same year as the seminal work of Boole -Boole [1847] that started mathematics as a theoretical discipline, Augustus De Morgan published a foundational work in logic [De Morgan, 1847]. In this book, De Morgan defines and describes symbolic mathematical logic. His work has become the foundation for our current mathematical system. One of the key components of this work examines the complements of intersections and unions [De Morgan, 1847, p. 69].

Theorem 2.5. *If A and B are subsets of the same universal set, then*

$$(A \cap B)^c = A^c \cup B^c \quad \textit{and} \quad (A \cup B)^c = A^c \cap B^c.$$

Proof. Because we are proving that two sets are equal we need to prove that the sets are subsets of each other. Here we will prove that $(A \cap B)^c = A^c \cup B^c$ and leave the other proof as an exercise.

(Proof that $(A \cap B)^c \subseteq A^c \cup B^c$.):

Let $x \in (A \cap B)^c$. So x is not in $A \cap B$. x is not in both A and B.

Case 1. If $x \in A$, then $x \notin B$. So $x \in B^c$.

Case 2. If $x \notin A$, then $x \in A^c$.

So either way, x is in A^c or x is in B^c ($x \in A^c \cup B^c$). Therefore, $(A \cap B)^c \subseteq A^c \cup B^c$.

(Proof that $A^c \cup B^c \subseteq (A \cap B)^c$.):

Let $x \in A^c \cup B^c$.

Case 1. $x \in A^c$. Then $x \notin A$. So $x \notin A \cap B$. So $x \in (A \cap B)^c$.

Case 2. $x \in B^c$. Then $x \notin B$. So $x \notin A \cap B$. So $x \in A \cap B)^c$.

Therefore, $A^c \cup B^c \subseteq (A \cap B)^c$

Therefore, $(A \cap B)^c = A^c \cup B^c$. □

It is also helpful to understand De Morgan's laws by looking at the corresponding Venn diagrams, shown in Figure 2.7.

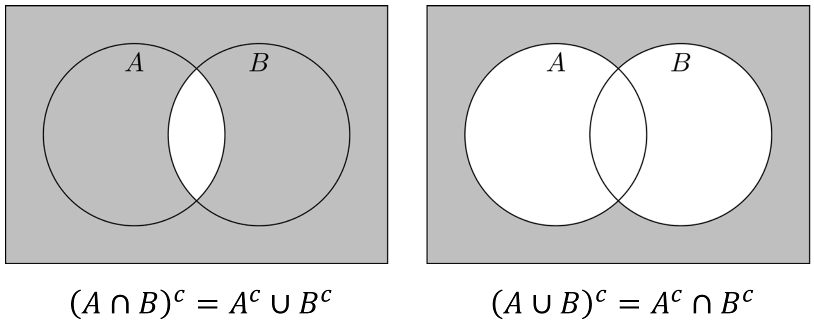


Figure 2.7: Venn diagrams for De Morgan's laws for pairs of sets

2.2.3 Cartesian Products

The previous sections have considered set operations between two sets that exist in the same universal set. Sets can also be combined to create new sets that exist in a universal set that differs from those of the sets used to create it. The collection of set operations that do this allows us to use sets to create multidimensional systems such as ordered pairs.

Definition 2.7. The **Cartesian Product** of two non-empty sets A and B is the set

$$A \times B = \{(a, b) | a \in A, b \in B\}$$

of all ordered pairs of elements where the first coordinate in the pair comes from the set A and the second coordinate is an element of B.

The Cartesian product, $A \times B$, is often read as "A cross B".

The most common Cartesian product in the secondary mathematics curriculum is real plane (often called the Cartesian plane or coordinate plane), $\mathbb{R} \times \mathbb{R}$, which is often denoted by

$$\mathbb{R}^2 := \{(x, y) | x, y \in \mathbb{R}\}.$$

Note that the symbol $:=$ is read as "defined to be equal to" and is a short hand for definitions in mathematics.

Related Content Standards

- (5.G.1) Use a pair of perpendicular number lines, called axes, to define a coordinate system, with the intersection of the lines (the origin) arranged to coincide with the 0 on each line and a given point in the plane located by using an ordered pair of numbers, called its coordinates. Understand that the first number indicates how far to travel from the origin in the direction of one axis, and the second number indicates how far to travel in the direction of the second axis, with the convention that the names of the two axes and the coordinates correspond (e.g., x-axis and x-coordinate, y-axis and y-coordinate).

The concept of the Cartesian product can be generalized to more than a pair of sets, for example

$$\mathbb{R}^3 := \{(x, y, z) | x, y, z \in \mathbb{R}\}$$

is the three dimensional Cartesian space where each coordinate is a real number.

Theorem 2.6. *It is sometimes helpful to summarize all of the properties of algebra on sets into a single location.*

Let all sets referred to below be subsets of a universal set U.

1. *(Commutative Laws) For all sets A and B,*

$$(a)\ A \cup B = B \cup A \quad \text{and} \quad (b)\ A \cap B = B \cap A$$

2. *(Associative Laws) For all sets A, B, and C,*

$$(a)\ (A \cup B) \cup C = A \cup (B \cup C) \quad \text{and} \quad (b)\ (A \cap B) \cap C = A \cap (B \cap C)$$

3. *(Distributive Laws) For all sets A, B, and C,*

$$(a)\ A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{and} \quad (b)\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

4. *(Identity Laws) For all sets A,*

$$(a)\ A \cup \emptyset = A \quad \text{and} \quad (b)\ A \cap U = A$$

5. *(Complement Laws)*

$$(a)\ A \cup A^c = U \quad \text{and} \quad (b)\ A \cap A^c = \emptyset$$

6. *(Double Complement Law) For all sets A,*

$$(A^c)^c = A$$

7. *(Idempotent Laws) For all sets A,*

$$(a)\ A \cup A = A \quad \text{and} \quad (b)\ A \cap A = A$$

8. *(Universal Bound Laws) For all sets A,*

$$(a)\ A \cup U = U \quad \text{and} \quad (b)\ A \cap \emptyset = \emptyset$$

9. *(De Morgan's Laws) For all sets A and B,*

$$(a)\ (A \cup B)^c = A^c \cap B^c \quad \text{and} \quad (b)\ (A \cap B)^c = A^c \cup B^c$$

10. *(Absorption Laws) For all sets A and B,*

$$(a)\ A \cup (A \cap B) = A \quad \text{and} \quad (b)\ A \cap (A \cup B) = A$$

11. *(Complements of U and $\emptyset$)*

$$(a)\ U^c = \emptyset \quad \text{and} \quad (b)\ \emptyset^c = U$$

12. *(Set Difference Law) For all sets A and B,*

$$A \setminus B = A \cap B^c$$

2.2.4 Exercises

1. Middle and high school students often struggle to remember the difference between union and intersection.

   a) Describe a memory trick to help students remember which symbol goes with which of the two operations.
   b) Review the definitions of the intersection and union of two sets. What key words separate the two definitions from each other?

c) Some students, when first learning the mathematical definition of union, think that the definition excludes objects that are in both sets. These students, when given two sets and asked to find $A \cup B$ will include the items that are in A only and B only. They exclude things in $A \cap B$. What might be the source of this misconception?

d) Define a non-numeric universe and two sets, A and B, in your universe such that $A \cap B \neq \emptyset$. Describe, using words, each of the following sets:

   i. $A \cup B$
   ii. $A \cap B$
   iii. A^C

2. Let $A = \{1, 3, 5, 7, 9\}$, $B = \{1, 2, 3, 4\}$, and $C = \{3, 6, 9\}$. List the elements of each of the specified sets.

   a. $A \cap B$
   b. $A \cup B$
   c. $A \cup C$
   d. $(A \cap B) \cup C$
   e. $A \cap (B \cup C)$
   f. $A \times B$
   g. $B \times (A \cap C)$

3. For this exercise, assume that $\mathbb{R}$ is the universal set. For any natural number, n, define $n\mathbb{Z} = \{nx | x \in \mathbb{Z}\}$. Answer the following as true or false. If false, explain why the statement is not true.

   a. $(2\mathbb{Z})^C = \{2x + 1 | x \in \mathbb{Z}\}$
   b. $\mathbb{R} \setminus \mathbb{Z} = \mathbb{Z}^C$
   c. $5\mathbb{Z} \cap \{2x + 1 | x \in \mathbb{Z}\} = 5\mathbb{Z}$
   d. $5\mathbb{Z} \cap 4\mathbb{Z} = 20\mathbb{Z}$
   e. $2\mathbb{Z} \setminus (4\mathbb{Z} \cup 6\mathbb{Z}) = \emptyset$
   f. $3\mathbb{Z} \setminus 2\mathbb{Z} = \{3(2x - 1) | x \in \mathbb{Z} \text{ and } x \geq 0\}$

4. Let A and B be sets. Prove that $A \subseteq A \cup B$.

5. Let A and B be sets. Prove that if $A \subseteq B$, then $A \cup B = B$.

6. Prove Theorem 2.3.

7. Prove Part 2 of Theorem 2.4.

8. Prove that for sets A and B, $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$

9. Write $(A \setminus B) \cup (B \setminus A)$ in terms of just unions, intersections, and complements, then simplify your expression.

10. Often when doing mathematics, the set you are working with or within is left unstated.

    a. Under what conditions is it important to be explicit with students about the set you are working within?
    b. What set do you assume you are working in when:
       1. You are figuring out how much something will cost?
       2. You are figuring out what proportion of a pizza to give everyone?
       3. You are determining what the temperature will be if it is predicted to drop 20 degrees overnight?

11. Construct an algebraic proof for the given statement. Cite a property from Theorem 2.6 for every step.

    For all sets A and B,
$$A \cup (B \setminus A) = A \cup B$$

2.3 Collections of Sets

Now that we know how to combine pairs of sets, we can inductively define unions and intersections for a finite or infinite number of sets.

When we are dealing with more than one or two related, but distinct sets, we often use another set as an **index set** in order to more easily describe and distinguish the sets in the collection. The most common indexing set used is the set of natural numbers, $\mathbb{N} = \{1, 2, 3, \ldots\}$. Then we use this indexing set as a label for the sets that we are considering. So for each natural number, there is a corresponding set, A_n, usually related to the value of n in some way.

Example 2.11. Let the natural numbers, $\mathbb{N} = \{1, 2, 3, \ldots\}$, be an indexing set. For each natural number, n, we define

$$A_n = [n, n+1),$$

which is an interval in the real numbers. In this situation, we have

$$A_1 = [1, 2),\ A_2 = [2, 3),\ A_3 = [3, 4),\ \ldots$$

In this case, the use of $\mathbb{N}$ tells us that we have an infinite set of sets. We use the indexing set to express the form of a general interval. When we assign an index value (such as $n = 3$), we are effectively identifying a particular interval in the sequence of intervals.

Using indexing sets, we can then define the union and intersection of a collection of sets.

Definition 2.8. Let S be an indexing set and let $\{A_i\}_{i\in S}$ be an indexed non-empty family of sets. Then we define the union and intersection of the family of sets as

$$\bigcup_{i\in S} A_i = \{x|x \in A_i \text{ for some } i \in S\} \quad \text{and} \quad \bigcap_{i\in S} A_i = \{x|x \in A_i \text{ for all } i \in S\}.$$

Example 2.12. Continuing with our previous example where $A_n = [n, n+1)$ for each $n \in \mathbb{N}$, we can write

$$\bigcup_{i\in\{1,2,3\}} A_i = [1, 4) \quad \text{and} \quad \bigcap_{i\in\{1,2,3\}} A_i = \emptyset$$

since the sets A_1, A_2, A_3, and A_4 do not overlap. If we extend this to the entire indexing set, then

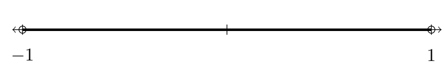
$$\bigcup_{i\in\mathbb{N}} A_i = [1, \infty) \quad \text{and} \quad \bigcap_{i\in\mathbb{N}} A_i = \emptyset.$$

An indexed collection of sets $\{A_i\}_{i\in S}$ is called **mutually disjoint** if, for any $i, j \in S$ with $i \neq j$, $A_i \cap A_j = \emptyset$. The sets in the previous example are mutually disjoint since $[n, n+1) \cap [m, m+1) = \emptyset$ if $m \neq n$.

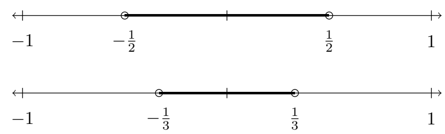
Example 2.13. For each positive integer n, ($n \in \mathbb{N}$), let

$$S_n = \left\{ x \in \mathbb{R} \middle| \frac{-1}{n} < x < \frac{1}{n} \right\}.$$

$$S_1 = (-1, 1)$$



$$S_2 = (\frac{-1}{2}, \frac{1}{2})$$



$$S_3 = (\frac{-1}{3}, \frac{1}{3})$$

Then we can see that for any $i < j$, we have that $S_i \cap S_j = S_j$ and $S_i \cup S_j = S_i$.

We can also take the union and intersections over the entire collection of sets.

$$\bigcup_{n\in\mathbb{N}} S_n = (-1, 1) \quad \text{and} \quad \bigcap_{n\in\mathbb{N}} S_n = \{0\}.$$

The above example is also an example of what is called a **nested** collection of sets.

Definition 2.9. An indexed collection of sets, $\{A_i\}_{i\in I}$, with an order on I is called **nested** if either

$$A_i \subseteq A_j \text{ whenever } i < j \quad \text{or} \quad A_i \supseteq A_j \text{ whenever } i < j.$$

When the indexing set is the natural numbers this is often denoted by

$$A_0 \subseteq A_1 \subseteq A_2 \subseteq \cdots \quad \text{or} \quad A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots .$$

Note: the student is strongly encourage to complete the related exercises to see addition examples of nested collections of sets. We can also see from the definitions that a collection of sets cannot be both nested and mutually disjoint.

Using the ideas of indexing sets and families of sets, we can generalize De Morgan's Laws to a general collection of sets, the proofs of which are very similar to the proof in the case of two sets.

Theorem 2.7. *Let I be an indexing set and let $\{A_i\}_{i\in I}$ be a collection of sets that are all subsets of the same universal set. Then*

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c \quad \textit{and} \quad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c$$

2.3.1 Exercises

1. For each of the following collections of sets:

$$\mathcal{A} = \left\{\left[\frac{1}{n}, n\right)\right\}_{n=2,3,4,\ldots}$$

$$\mathcal{B} = \{(n, \infty)\}_{n=0,1,2,3,4,5,\ldots}$$

$$\mathcal{C} = \{[-n, n]\}_{n=0,1,2,3,4,5,\ldots}$$

$$\mathcal{D} = \{[x, x+1)\}_{x\in\mathbb{R}}$$

$$\mathcal{E} = \{\{z \in \mathbb{C} | |z| = r\}\}_{r\in\mathbb{R}^+}$$

$$\mathcal{F} = \{\{n \in \mathbb{Z} | n = 3k + j \text{ for some } k \in \mathbb{Z}\}\}_{j=0,1,2}$$

a. Determine if the sets are mutually disjoint
b. Determine if the collection is nested
c. Find the union of the collection

Chapter 3

Equality, Order, and Equivalence

The chapter title may look like just a list, but the notions of equality, order, and equivalence are a critically important trio of related but distinct ideas in mathematics. Together they help us define when two things are the "same." Instruction in these ideas starts on almost the first day of school when students first learn how to count. Students continue study and expand these ideas in their study of mathematics: many of the theorems of mathematics labeled as "Fundamental" are statements of equivalence. These include the Fundamental Theorem of Arithmetic that every integer can be uniquely factored into primes, the Fundamental Theorem of Calculus describing the relationship between antiderivatives and infinite sums, and isomorphism theorems of abstract algebra and homeomorphisms of topology.

3.1 Partitions and Equivalence Relations

Figure 3.1: Sorting Candy

As a way to better understand the world, humans tend to sort information into categories or orderings. For instance, when some people see a pile of various types of M&M'sTM they have the urge to sort them in various ways. Some people sort the candies according to their color. Others may sort them according to their type (plain, peanut, etc.). This process of sorting is a type of "partitioning" the items into distinct categories. With this categorization we want to make sure that each candy is sorted into one, and only one, type. This type of sorting can be formalized with the following definition.

3.1.1 Partitions

Definition 3.1. A **partition** of a set A is a finite or infinite collection of non-empty, mutually-disjoint subsets whose union is A.

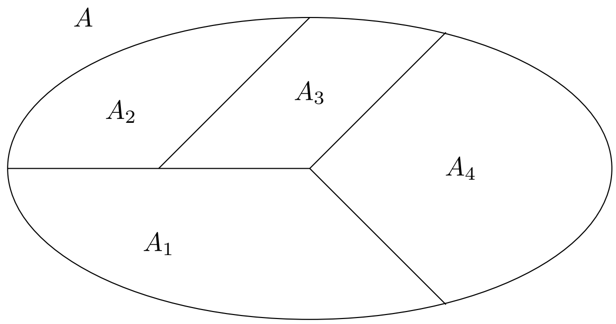


Figure 3.2: Partition of Sets

Thus, a partition of a set is a collection of subsets of the original set with several key properties.

1. The subsets are non-empty, meaning that each subset in the partition contains at least one element. This allows us to restrict the partition to a collection of non-trivial subsets and means partitions have a type of uniqueness.
2. The subsets in the partition are also mutually-disjoint which means that no element of the original set is in more than one of the subsets in the partition.
3. Lastly, since the union of the subsets that form the partition is A, we have that every element in A is in one of these subsets.

Thus, each component of the definition of a partition is essential to defining a way to split a set into a set of categories that we find useful.

Example 3.1. If we have a set

$$A = \{a, 2, \alpha, \text{apple}, \text{box}, \beta, 1, 3, \gamma, \text{cow}, c, b\}$$

there are some natural ways to partition the set. Here are three possible partitions.

$$A_1 = \{a, b, c\},\ A_2 = \{1, 2, 3\},\ A_3 = \{\alpha, \beta, \gamma\},\ A_4 = \{\text{apple}, \text{box}, \text{cow}\}$$
$$A_1 = \{\text{apple}, \text{box}, \text{cow}\},\ A_2 = \{a, b, c, 1, 2, 3, \alpha, \beta, \gamma\}$$
$$A_1 = \{a, 1, \alpha, \text{apple}\},\ A_2 = \{b, 2, \beta, \text{box}\},\ A_3 = \{c, 3, \gamma, \text{cow}\}$$

However,

$$A_1 = \{a, b, c, \text{apple}, \text{box}, \text{cow}\}, A_2 = \{\alpha, \beta, \gamma, \text{apple}, \text{box}, \text{cow}\}, A_3 = \{1, 2, 3, \text{apple}, \text{box}, \text{cow}\}$$

is not a partition of the set because A_1, A_2, and A_3 are not mutually disjoint.

There are many other ways to partition the set, but we do not have the space to write out all of the possible partitions of this set.

Example 3.2. Early in the K-12 mathematics curriculum we ask students to determine which integers are even and which are odd. In this situation, the set A is the set of integers ($\mathbb{Z}$), while the subsets that form the partition are the subset of even numbers, ($2\mathbb{Z}$), and odd numbers, ($2\mathbb{Z} + 1$).

Example 3.3. If we let the set B be defined by the following list of elements,

$$B = \left\{ \begin{array}{ccccc} \text{jar,} & \text{jellybean,} & \text{applesauce,} & \text{ape,} & \text{bear,} \\ \text{ball,} & \text{beanbag,} & \text{bag,} & \text{arm,} & \text{skate,} \\ \text{bike,} & \text{rock,} & \text{rowboat,} & \text{jigsaw,} & \text{saw} \end{array} \right\}$$

take some time to create various partitions of the set B, verifying the various properties in the definition of a partition.

In this text, we denote the partition $\mathcal{P}$ of a set A as the union of subsets of A over some indexing set, I. For example, if the partition is formed using two subsets of A, then the indexing set could be $I = \{1, 2\}$. The indexing set can also be infinite. For instance,

$$\mathcal{P} = \{A_n\}_{n\in\mathbb{Z}}, \text{ where } A_n = \{x \in \mathbb{R} | n \leq x < n+1\}$$

is a partition of the real numbers with an infinite number of sets.

It is important to note that there are usually many ways to partition a set, but the definition of partition does not impose any constraints on the uniformity or order of the subsets, only that the way we choose our subsets meets the three conditions described previously.

With our previous discussion and examples in mind, we can translate our verbal definition of a partition into a symbolic one. In particular: if $\mathcal{P} = \{U_i\}_{i\in I}$ is a partition of a set A, where I is some indexing set that is either finite or infinite, we have the following properties

- the elements of the partition are mutually disjoint:
$$U_i \bigcap U_j = \emptyset \quad \text{if } i \neq j,$$
- the union of the elements of the partition is the entire set:
$$\bigcup_{i\in I} U_i = A,$$
- the elements of the partition are non-empty:
$$U_i \neq \emptyset \text{ for all } i \in I.$$

3.1.2 Relations

Up until this point we have focused on building a framework for understanding the notion of partition. In order to understand the main topics of this chapter, we also need to develop the idea of a relation. While we give the general definition of a relation and a few examples, we will focus on those relations that arise from partitions in this section. We will explore other types of relations in later sections.

Definition 3.2. Let A and B be sets. We define a relation R from A to B as a subset of $A \times B$. We say that for elements $a \in A$ and $b \in B$ that a is related to b, (aRb), if and only if $(a, b) \in R$. If the two sets are the same set, A, then we say that R is a relation on A.

Note that relations can be written in many different ways, with the most common being similar to $a = b$, $a < b$, $a\tilde{b}$.

Here are some examples of relations in the K-12 curriculum.

- Let A and B be the set of integers. We could say that
$$R = \{(a, b) \in A \times B | a < b\},$$
which defines the 'less than' relation of $aRb \leftrightarrow a < b$.

- Let A be the real numbers and B be the non-negative real numbers. We can let

  $$R = \{(x, y) \in A \times B | y = x^2\}.$$

  This is the graph of the function $f(x) = x^2$.

- Let A and B be the set of integers. We can define

  $$R = \{(m, n) | m - n = 2k, \text{ for some integer } k\}.$$

  This relation is one where are the even numbers are related to each other and all of the odd numbers are related to each other.

3.1.3 Equivalence Relations Induced by Partitions

For the remainder of this section we will study a specific type of relation that is created by a partition, $\mathcal{P}$, of a set, A. If $\mathcal{P} = \{U_i\}_{i \in I}$ is a partition of A, we can write A as the union of a collection of non-empty, mutually disjoint subsets:

$$U_i \cap U_j = \emptyset \text{ if } i \neq j, \text{ for each } i \in I, U_i \neq \emptyset, \text{ and } A = \bigcup_{i \in I} U_i.$$

Consider one of the individual subsets in the partition, say U_i. Then the elements of U_i, by virtue of the fact they are in the same subset, have a relationship to each other. This idea extends to the whole partition. That is, the partition induces a relation on A in the following way. For each $a, b \in A$ we say that aRb if, and only if, there is a set U_i in the partition P such that both a and b are in U_i.

The relations we create through a partition behave in ways that are of general interest in the study of mathematics.

If we choose a generic element a in A, the property that $\bigcup_{i \in I} U_i = A$ tells us that there exists $i \in I$ such that $a \in U_i$. So for every element $a \in A$, we have that a is related to itself, aRa. We will call this property of the relation the **reflexive property**.

For the second property, if we have $a, b \in A$ such that aRb, then there must be a $U_i \in P$ with $a, b \in U_i$. This also means that bRa. When you have the property that aRb implies that bRa, we say that the relation has the **symmetric property**.

The third property of relations that we define derives from the property of partitions that

$$U_i \bigcap U_j = \emptyset \Leftrightarrow i \neq j.$$

If $a, b, c \in A$ such that aRb and bRc, then we have that there are sets U_i and U_j in the partition such that a and b are both in U_i and b and c are both in U_j. Since b is in both U_i and U_j, $U_i \cap U_j \neq \emptyset$ and so $i = j$. This would then imply that aRc. This property that the combination of aRb and bRc implies that aRc is called the **transitive property** of relations.

Not all relations have these three properties of reflexive, symmetric, and transitive. Those relations that have all three properties are called **equivalence relations**.

We have therefore proven that a partition of a set induces an equivalence relation on that set.

3.1.4 Partitions Induced by Equivalence Relations

The concept of equivalence relation is one that transcends much of mathematics and we often consider things as being the "same" with respect to some property if there is an equivalence relation under which they are related. One location that this appears in the elementary school curriculum is in regards to rational numbers.

We think of $\frac{1}{2}$ and $\frac{2}{4}$ as being the same rational number, but written in a different way. Along these lines we say that two rational expressions $\frac{a}{b}$ and $\frac{c}{d}$ are equivalent if and only if $ad = bc$. This creates and equivalence relation between these rational expressions that are used to create the rational numbers.

Similarly, if a relation, R, on a set, A, is an equivalence relation, then one can generate an induced partition, $\mathcal{P}$, as follows. For each $a \in A$ we define the set

$$[a]_R := \{b \in A | aRb\}.$$

We call this set, $[a]_R$, the equivalency class of a under the relation R.

Since R is an equivalence relation, it is reflexive and so $a \in [a]_R$ causing $[a]_R \neq \emptyset$. Since $a \in [a]_R$, we have that

$$\bigcup_{a \in A} [a]_R = A.$$

Thus to show that $\mathcal{P} = \{[a]_R | a \in A\}$ is a partition of A, we need to show that for $a, b \in A$ that $[a]_R = [b]_R$ or $[a]_R \bigcap [b]_R = \emptyset$.

If $[a]_R \bigcap [b]_R \neq \emptyset$, then we need to prove that two sets to be equal.

Let us assume that there exists a $c \in [a]_R \bigcap [b]_R$.

Let $d \in [a]_R$. So aRd. c is also in $[a]_R$ and so aRc. Since R is symmetric and aRd, dRa. We now have that dRc using transitive property. c is also in $[b]_R$ and so bRc. Using the symmetric property, we have cRb. Combining this with dRc and using the transitive property we have dRb. Using the symmetric property we have bRd. So $d \in [b]_R$. Therefore, $[a]_R \subseteq [b]_R$.

Using very similar arguments we can prove that $[b]_R \subseteq [a]_R$.

So if the intersection is nonempty, the two sets are the same set.

We have thus proven the following theorem showing the connection between equivalence classes and partitions.

Theorem 3.1. *If R is an equivalence relation on a set A, then*

$$P = \{[a]_R | a \in A\}$$

is a partition of A.

Similarly, if $P = \{U_i | i \in I\}$ is a partition on a set A, then the relation R defined by

$$aRb \Leftrightarrow \textit{ there exists } i \in I \textit{ such that } a, b \in U_i$$

is an equivalence relation.

In summary, we have defined a relation R from A to B as a subset of $A \times B$, and that a is related to b, (aRb), if and only if $(a, b) \in R$. If the two sets are the same set, A, then we say that R is a relation on A.

We also defined the following three properties of a relation, R, on a set A:

- **Reflexive.** The relation R is reflexive if aRa for all $a \in A$.
- **Symmetric.** The relation R is symmetric if aRb implies that bRa for all $a, b \in A$.
- **Transitive.** The relation R is transitive if aRb and bRc imply that aRc for $a, b, c \in A$.

Note: to prove that a relation satisfies one of these properties, you need to show that the statement is true for all possible elements. To prove that a property is not satisfied, a single counter example is sufficient.

A relation is an equivalence relation if, and only if, it satisfies all three of these properties.

3.1.5 Ordered Sets and Relations

Another type of relation is one that creates an order, or ranking, of the set. A common mathematical topic in this regards are the concepts of “less than” and “less than or equal to”.

We will first explore the relation of “less than”. Since a number cannot be less than itself we see that the relation is not reflexive. We also know that $a < b$ and $b < a$ cannot both be true and so the relation is not symmetric. However, if $a < b$ and $b < c$, we do have that $a < c$. So this relation of “less than” is transitive. Thus the familiar relation of “less than” is one which is transitive but not reflexive or symmetric.

When we expand the relation to that of “less than or equal to”, we add that $a \leq a$, resulting in the relation being reflexive. So this is a common relation that is reflexive and transitive, but not symmetric. For example, $1 \leq 2$, but 2 is not less than or equal to 1.

Definition 3.3. We define an **order** on a set, A, to be a relation, denoted by $<$, with the following properties.

1. If x and y are two elements of A, then one, and only one, of the statements
$$x < y, \quad x = y, \quad , y < x$$
is true.
2. If $x, y, z \in A$, and if $x < y$ and $y < z$, then $x < z$.

If a set has an order defined on the set, then we say that it is an **ordered set**.

Such an order forms a relation that is transitive, but not symmetric or reflexive. As a convenient notation, we will write $x \leq y$ as the negation of $y < x$. We then see that it is equivalent to the statement that ($x < y$ or $x = y$).

Definition 3.4 (Well-Ordering Property)**.** A set, A, together with an order, $<$, satisfies the well-ordering property if every non-empty subset of A has a least element. Equivalently, if $S \subseteq A$ with $S \neq \emptyset$, then there exists an element $s_0 \in S$ such that $s_0 \leq s$ for all $s \in S$.

We will see in Chapter 4 that the natural numbers, $\{0, 1, 2, 3, \ldots\}$, satisfy the well-ordering property, but the real numbers do not. For instance, $\{x \in \mathbb{R} | x > 1\}$ does not have a least element. This means that in some way, a set needs to be discrete in order to satisfy a well-ordering property.

For sets that are not discrete, we would like to have a similar property.

Definition 3.5. Let A be an ordered set and let $S \subseteq A$. If there exists an $\alpha \in A$ such that $x \leq \alpha$ for all $x \in S$, then we say that S is **bounded above** and that such an α is an **upper bound** of S.

We can similarly define **bounded below** and **lower bound** using opposite inequalities.

Definition 3.6 (Least-Upper-Bound Property)**.** An ordered set A is said to have the **least-upper-bound property** if for every non-empty subset $S \subseteq A$ that is bounded above, there exists an element $s_0 \in S$ such that s_0 is an upper bound of S and if α is an upper bound of S, then $s_0 \leq \alpha$. Such an s_0 is called the **least-upper-bound** of S.

We will see in Chapter 4 that the real numbers satisfy the least-upper-bound property, while the rational numbers do not.

3.1.6 Exercises

1. For each of the following relations, determine if the relation is reflexive, symmetric, and/or transitive.

   a. Let S be a relation defined on $\mathbb{R}$ by
   $$xSy \Leftrightarrow x^2 = y^2.$$
   b. Let T be a relation defined on $\mathbb{R}$ by
   $$xTy \Leftrightarrow x < y^2.$$
   c. Let U be a relation defined on $\mathbb{Z}$ by
   $$mUn \Leftrightarrow mn > 0.$$
   d. Let V be a relation defined on $\mathbb{R}^2$ by
   $$(x,y)V(w,z) \Leftrightarrow x^2 + y^2 \leq w^2 + z^2.$$

2. Let $S = \{a, b\}$. List out all of the possible relations on S and determine which relations are reflexive, symmetric, and/or transitive.

3. Define a relation $\sim$ on $\mathbb{R} \times (\mathbb{R} - \{0\})$ by
   $$(x,y) \sim (z,w) \Leftrightarrow xw = yz.$$
   a. Show that this relation is an equivalence relation.
   b. Give a description of the partition of $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$ induced by this equivalence relation.
   c. Why do we not allow 0 in the second part of the ordered pairs?

4. Determine which of the following collections of sets form a partition of $\mathbb{R} \times \mathbb{R}$. For those that do form a partition, define the equivalence relation (you do not need to prove each of the properties). For those that do not form a partition, what is one property that fails.

   a. For $c \in \mathbb{R}$, let $h_c = \{(x,y) \in \mathbb{R}^2 | y = c\}$, and let $\mathcal{C} = \{h_c\}_{c \in \mathbb{R}}$.
   b. For $m \in \mathbb{R}$, let $s_m = \{(x,y) \in \mathbb{R}^2 | y = mx\}$, and let $\mathcal{D} = \{s_m\}_{m \in \mathbb{R}}$.
   c. For $r \in [0,\infty)$, let $C_r = \{(x,y) \in \mathbb{R}^2 | x^2 + y^2 = r^2\}$, and let $\mathcal{E} = \{C_r\}_{r \in [0,\infty)}$.

5. (For students who have knowledge of linear algebra) Let $M_{2,2}$ be the set of 2×2 matrices with real coefficients. For each of the following, prove that it is an equivalence relation, or show that one of the properties of equivalence relations does not hold.

   a. For matrices A and B in $M_{2,2}$, let $A \sim B$ if and only if $\det(A) = \det(B)$.
   b. For matrices A and B in $M_{2,2}$, let $A \sim B$ if and only if $\mathrm{tr}(A) = \mathrm{tr}(B)$.
   c. For matrices A and B in $M_{2,2}$, let $A \sim B$ if and only if the eigenvalues of A are the same as the eigenvalues of B.

3.2 Equality and Equivalence in the K-12 Curriculum

We can see from the discussion in the previous two sections that the concept of two things being equivalent mathematically is a very difficult and abstract concept. It is no surprise then that students of all grade levels have many different conceptions and misconceptions revolving around this concept of equality and equivalence.

Since all of the number systems that arise in the elementary and early secondary curriculum have an inherent order described by the $\leq$ and $<$ symbols, the concept of equality is often described in terms of a balanced scale. Another conception around the equal sign is that two sets of objects have the same number of items,

1. Solve the following:
a) $5 + 2 =$
b) $3 + 4 =$

Figure 3.3: Sample problems

determined by counting the number of objects. This is likely since the concept of equality arises in the first grade where students are only working with whole numbers.

One of the common misconceptions surrounding the equal sign is that the symbol implies an action of doing some type of computation. For instance, students usually see the equal sign in the context of the type of problems seen in Figure 3.3.

These type of problems promote an operational understanding of the equal sign, rather than a relational or substitutional understanding that is essential for later work [Gilmore et al., 2018, pp. 145-150]. Because secondary mathematics students maintain many misconceptions about the concept of equality and the meaning of the equal sign, we will take some time to reflect on some key content standards from first grade that have a very strong influence on student success of learning mathematics in the middle school.

Related Content Standards

- (1.OA.7) Understand the meaning of the equal sign, and determine if equations involving addition and subtraction are true or false. For example, which of the following equations are true and which are false? $6 = 6$, $7 = 8 - 1$, $5 + 2 = 2 + 5$, $4 + 1 = 5 + 2$.
- (1.OA.8) Determine the unknown whole number in an addition or subtraction equation relating three whole numbers. For example, determine the unknown number that makes the equation true in each of the equations $8+? = 11$, $5 =? - 3$, $6 + 6 =?$.

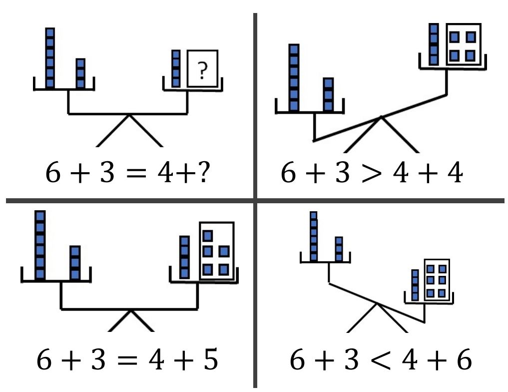


Figure 3.4: Pan balances

We should notice that in standard (1.OA.D.8) that the unknown should not always be at the end of a string of mathematical symbols and is based on educational research that compares students from different cultures and their understanding of the equal sign [Gilmore et al., 2018, pp. 148-150]. Instead it should be at many different locations in order to make sure that students do not internalize the misconception that the equal sign means to compute. They should instead develop the concept of the equal sign representing two things that have the same value in some way. It is this understanding of two things looking different but representing the same underlying value that causes the equal sign to represent an equivalence.

Another method that may possibly contribute to a more relational understanding of the equal sign is the introduction of the equal sign alongside the greater than (>) and less than (<) symbols to emphasize the role of the equal sign in the comparison of the size of numbers. This comparison concept can be further reinforced with a pan balance (See Figure 3.4).

This concept of a pan-balance can give students a concrete idea to attach to the meaning of the equal and inequality signs. However, like all good analogies, the pan balance is challenging to use with negative numbers and breaks down if you are comparing numbers in an unordered number system such as the complex numbers.

3.2.1 Exercises

1. A student is given the following task.

   > Sophia goes to the store with \$5.00 and buys two packs of gum for \$1.20. She then gives \$2.00 to her friend to buy a drink. How much money does she have left?

   The student then writes the following on their paper.

   $$5.00 - 1.20 = 3.80 - 2.00 = 1.80, \text{ so Sophia has \$1.80 left.}$$

   a. What ideas from this section regarding the equal sign and equivalence would you use to provide feedback to this student?

   b. What are other points of feedback that you would provide to this student?

2. A student asks you what the equal sign means in the context of $\frac{1}{2} = \frac{2}{4}$ since those symbols are not the same. How would you respond to this student? How would it change depending on the grade level of the student?

3. What are the meanings of the equal signs in the task?

   > Evaluate the function $f(x) = 3x - 4$ for $x = 2$.

3.3 Expressions, Equations, and Inequalities

A **numerical expression** is a meaningful string of numbers, operation symbols, and/or grouping symbols. In early elementary grades, examples of numerical expressions include

$$2 + 6 + 4, \quad 13 - 3 + 2, \quad \text{or} \quad (4 + 6) + 3.$$

In upper elementary grades (3-6), students are expected to incorporate rational numbers and the operations of multiplication, division, and exponents in order to be comfortable working with such expressions as

$$3 \times (10 + 4), \quad \frac{2}{3} \times (12 + 8), \quad \text{or} \quad 2^3 + \frac{3 + 2}{7}.$$

In the transition from elementary school to secondary schools, students learn to add **variables**, symbols or letters that stand for any number within a specified range of numbers, to these expressions and thus work with **algebraic expressions** such as

$$100 - 10 \cdot (P - 15)^3, \quad 3xy + 5(x - 4)^2 - 7y^2, \quad 6 \cdot (4x + 3y), \quad \text{or} \quad \frac{3x + 2}{2x - 4}.$$

Related Content Standards

- (6.EE.2) Write, read, and evaluate expressions in which letters stand for numbers.

a. Write expressions that record operations with numbers and with letters standing for numbers.
b. Identify parts of an expression using mathematical terms (sum, term, product, factor, quotient, coefficient); view one or more parts of an expression as a single entity.
c. Evaluate expressions at specific values of their variables. Include expressions that arise from formulas used in real-world problems. Perform arithmetic operations, including those involving whole-number exponents, in the conventional order when there are no parentheses to specify a particular order (Order of Operations).

In more advanced courses in secondary school mathematics, algebraic expressions can also include functions giving rise to expressions such as

$$f(x-h)+k, \quad \sin(3x)-2, \quad \text{or} \quad \ln(x+3).$$

The equivalence of numerical expressions is a key concept in the early elementary grades where students learn to compose and decompose numbers in different ways as they add and subtract numbers using different algorithms. In sixth grade, students are asked to determine the equivalence of two mathematical expressions. This definition that two algebraic expressions are equivalent if they generate the same number regardless of which number is substituted for the variable is one of the first places where students are pushed to move to the abstract realm of equivalence relations. We can verify that this definition of equivalent expressions is an equivalence relation using the fact that equality on the real numbers is an equivalence relation.

Related Content Standards

- (6.EE.3) Apply the properties of operations to generate equivalent expressions. %{*For example, apply the distributive property to the expression* $3(2+x)$ *to produce the equivalent expression* $6+3x$*; apply the distributive property to the expression* $24x+18y$ *to produce the equivalent expression* $6(4x+3y)$*; apply properties of operations to* $y+y+y$ *to produce the equivalent expression* $3y$*.}*
- *(6.EE.4) Identify when two expressions are equivalent (i.e., when the two expressions name the same number regardless of which value is substituted into them).*
- *(6.EE.5) Understand solving an equation or inequality as a process of answering a question: which values from a specified set, if any, make the equation or inequality true? Use substitution to determine whether a given number in a specified set makes an equation or inequality true.*

While substituting in values to determine if two expressions are equivalent is the definition of the equivalence, it is almost always impossible to substitute in all possible values for a variable, as there are infinitely many possibilities in most situations. Instead, we determine the equivalency of two expressions by applying various properties of the operations. For example, the distributive property shows us that $3(x+y)$ is equivalent to $3x+3y$. However, students have to be careful to consider the set of possible numbers for an expression when determining the equivalence of $\sqrt{x^2}$ and x, or $\frac{3x^2}{x}$ and $3x$.

An **equation** is a statement that a number or expression is equivalent to a different number or expression. An equation that is true for all values of the variable is called an **identity**. The associative, commutative, and distributive properties of the number systems

$$a+(b+c)=(a+b)+c$$
$$a+b=b+a$$
$$a\times(b+c)=(a\times b)+(a\times c)$$

are all common identities. Some other common identities in this regard are the trigonometric identity

$$\cos^2(\theta)+\sin^2(\theta)=1$$
$$\sin(\alpha\pm\beta)=\sin(\alpha)\cos(\beta)\pm\cos(\alpha)\sin(\beta)$$
$$\cos(2\theta)=\cos^2(\theta)-\sin^2(\theta)$$
$$\sin(2\theta)=2\sin(\theta)\cos(\theta)$$

that are used to simplify trigonometric expressions.

An equation that describes how multiple quantities vary together is usually called a **formula**. These can include the area formula for rectangles defined by Area = width × height or Boyle's law where pressure (P) is inversely proportional to volume (V), i.e. $PV = k$ for some constant k.

While the identity equations hold true irrespective of the values of the variables, it is often the case that equations hold for only some of the possible values for the variables, or possibly even none. We can then **solve** such equations by determining which values, called **solutions to the equation**, for the variables cause the equation to be a true statement.

Similar to equations, **inequalities** are statements that a number or expression is related to a different number or expression in a certain order relation. Some of these inequalities are identities or formulas, such as $x \leq |x|, x \in \mathbb{R}$, while some have a solution set, such as $x^2 + y^2 < 1$.

Definition 3.7. Two equations (inequalities) are equivalent if they are true statements for the same values of the variables.

In the process of finding solutions to equations or inequalities, it is almost always helpful to rewrite an equation or inequality as an equivalent equation or inequality. For instance in an effort to find the solutions to

$$\frac{x^2 - 10x + 21}{3x - 12} = \frac{x - 5}{x - 4}$$

we want to find an expression that is equivalent to the left side of the equation. Such an expression could be

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

with the equivalency of the expressions verified using the distributive property. Since these two expressions are equivalent, they have the same values for every value of the variable. This means that the original equation is a true statement if, and only if,

$$\frac{(x-3)(x-7)}{3(x-4)} = \frac{x-5}{x-4}$$

is a true statement. We also know that if two numbers are the same, then 3 times each of the numbers are also the same. So the original statement is true if, and only if, the equivalent equation

$$\frac{(x-3)(x-7)}{x-4} = \frac{3(x-5)}{x-4}$$

is true. If $x = 4$, then the statement would have no meaning, as division by zero has no meaning. This means that we can eliminate 4 from the set of possible solutions. With $x \neq 4$ we can multiply both sides of the equation by $(x - 4)$ to generate the equivalent equation,

$$(x-3)(x-7) = 3(x-5) \quad \text{and} \quad x \neq 4.$$

We can use the distributive property again to create an equivalent equation of

$$x^2 - 10x + 21 = 3x - 15 \quad \text{and} \quad x \neq 4.$$

Since we can also add the same expression, $-3x + 15$, to both sides of the equation and create an equivalent equation, we see that the original equation is equivalent to

$$x^2 - 13x + 36 = 0 \quad \text{and} \quad x \neq 4.$$

Using the distributive property, we see that this equation is equivalent to

$$(x-4)(x-9) = 0 \quad \text{and} \quad x \neq 4.$$

So the original equation is a true statement if, and only if, $(x-4)(x-9) = 0 \quad \text{and} \quad x \neq 4$ is a true statement. This only occurs when the variable, x, is equal to 9.

Through this process we can see how equivalence of equations is essential in the process of finding solutions to equations. This means that as students make the transition from elementary to secondary, it is important for teachers to be understanding of the challenge of thinking about equivalent expressions and to be explicit about this relationship.

3.3.1 Exercises

1. How would you use the information from this section to respond to the following student questions? Give an answer in the context of your expected student population.

    a. What is the difference between an expression and equation?
    b. Why can we not solve an expression?
    c. What does it mean to "solve an equation"?
    d. What makes two expressions equivalent to each other?
    e. What is the difference between how the equality sign is used when solving an equation versus when working with expressions?

2. Prove that the definition of equivalent equations forms an equivalency relation.

Chapter 4

Number Systems

K-12 mathematics is rich in the study of number systems, with particular focus on the Natural Numbers, Integers, Rational Numbers, Real Numbers, and Complex Numbers. In this chapter we use the properties of set theory and equivalence classes discussed in Chapters 2 and 3 to construct these numbers systems. We follow this discussion with a study of some of the properties of number systems, connecting them to the K-12 curriculum. Lastly, we examine different representations of these number systems helpful for developing numerical fluency and conceptual understanding of the numerical operations.

This process of constructing the number systems is very abstract and challenging for those who have not previously studied these constructions. The process is included because it helps teachers to better understand the abstract nature of the number systems and some of the challenges of their students. While working with many of these number systems has become second nature for math teachers, students struggle much more than new teachers expect. So a deeper knowledge of the properties of the number system helps in this teaching situation.

$$\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R} \subseteq \mathbb{C}$$

As we are working with the various number systems, we will use varying notation for our operations. With all of the constructions, subtraction and division are not defined. We instead define the additive inverse and multiplicative inverse and so we can think of subtraction $a - b$ as $a + (-b)$ and division $a \div b$ as $a \times \frac{1}{b}$. We will also use various symbols to represent multiplication so that $a \times b$, $a \cdot b$, and ab all represent the same thing.

4.1 Natural Numbers and Integers

As part of the effort to axiomatize (reduce to a system of basic truths or axioms) mathematics in the late 19^{th} and early 20^{th} centuries, individuals such as Richard Dedekind (1831-1916), Georg Cantor (1845-1918), Giuseppe Peano (1858-1932), Alfred Whitehead (1861-1947), David Hilbert (1862-1943), Ernst Zermelo (1871-1953), Bertrand Russell (1872-1970), and Abraham Fraenkel (1891-1965) followed a program to start with the most basic of assumptions (axioms) and build a solid foundation for the field of mathematics. One of the greatest works in this program is the *Principia Mathematica* by Whitehead and Russell [1910; 1912; 1913] that built such a foundation on logic and set theory that it took until page 86 of the second volume to prove that $1 + 1 = 2$. While we will not go to the full extent of abstraction, we will build up the natural numbers and the integers from the axioms of Zermelo-Fraenkel set theory, together with the axiom of choice (ZFC axioms).

4.1.1 Natural Numbers

We will study the concepts of counting and related ideas of the natural numbers in Chapter 5, but initially we will use John von Neumann's [1923] definition for the natural numbers which are defined inductively by defining the symbols

$$0 := \emptyset, \quad 1 := \{\emptyset\}, \quad 2 := \{\emptyset, \{\emptyset\}\}, \quad 3 := \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\},$$

and so on. So for any natural number n, the next natural number is defined inductively as

$$n + 1 := S(n) := n \cup \{n\},$$

where you think of $S(n)$ as the successor of n. We sometimes define to natural number prior to n as n', so that $n = S(n')$.

This definition of the natural numbers is highly dependent upon the axiom of infinity from the ZFC axioms. In the formal language of the ZFC axioms, the axiom of infinity is stated as

$$\exists I(\emptyset \in I \wedge \forall x \in I((x \cup \{x\}) \in I),$$

where the symbol $\forall$ can be read as "for all". Or equivalently, there is a set I that contains the empty set as an element and that if $x \in I$, then $x \cup \{x\}$ is also in I. The set I is a larger set than $\mathbb{N}$, but allows for the existence of the natural numbers.

With our definition of the natural numbers, we also have that $S(n) \neq n$,

$$S(n) = S(m) \Leftrightarrow n = m,$$

and we can create an order on the set of symbols $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ by defining that

$$n \leq m \iff n \subseteq m.$$

One can prove that this set of symbols satisfies the well-ordering property (3.4) using the ZFC axioms, which enables us to prove the principle of mathematical induction creating the mathematical machinery necessary for the study of the natural numbers.

Theorem 4.1 (Principle of Mathematical Induction)**.** *Suppose that S is a subset of $\mathbb{N}$ such that $0 \in S$, and for all $n \in \mathbb{N}$, $n \in S$ implies that $S(n) \in S$.*

Then S must be the set of natural numbers.

We will leave the detailed proof for a text on symbolic logic. However, the application of this theorem is that if we have a mathematical statement that varies along the natural numbers, $P(n)$. If the statement $P(0)$ is true, and if $P(k)$ being true implies that $P(k+1)$ is true, then $P(n)$ must be true for all natural numbers.

We inductively define the operation of addition on these symbols so that for any integer a,

$$\begin{aligned} a + 0 &:= a, \\ a + 1 &:= a + S(0) = S(a + 0), \\ a + 2 &:= a + S(1) = S(a + 1), \\ &\vdots \\ a + S(b) &:= S(a + b) \end{aligned}$$

We similarly define multiplication inductively by

$$a \times 0 := 0, \quad a \times 1 := (a \times 0) + a = a, \quad \text{and} \quad a \times S(b) := (a \times b) + a.$$

With these operations we have the following properties of the natural numbers.

Theorem 4.2. *Let a, b, and c be natural numbers. Then*

- $a + b$ *and* $a \times b$ *are natural numbers (closure)*
- $a + (b + c) = (a + b) + c$ *and* $a \times (b \times c) = (a \times b) \times c$ *(associative property)*
- $a + b = b + a$ *and* $a \times b = b \times a$ *(commutative property)*
- $a + 0 = a$ *and* $a \times 1 = a$ *(additive and multiplicative identities)*
- $a \times (b + c) = (a \times b) + (a \times c)$ *(distributive property)*
- *If* $a \times b = 0$*, then* $a = 0$ *or* $b = 0$ *(or both). (no non-zero divisors)*

In order to get a flavor for the techniques involved in proving these properties, we will prove the associative property of addition for the natural numbers. The proofs of the remainder of the statements will be left to the reader to work through.

Proof. Let $a, b \in \mathbb{N}$. Then, for each $c \in \mathbb{N}$, we let $P(c)$ be the statement that

$$a + (b + c) = (a + b) + c.$$

If $c = 0$, then $b + c = b$ and $(a + b) + c = a + b$, and so

$$a + (b + c) = a + b = (a + b) + c.$$

This implies that $P(0)$ is true.

Assume by that for some value of $c' \in \mathbb{N}$ that $P(c')$ is true, i.e. that $a + (b + c') = (a + b) + c'$ ($c' = 0$ is true and so such a c' exists). Then for $c = S(c')$,

$$\begin{aligned} a + (b + c) &= a + (b + S(c')) = a + S(b + c') \\ &= S(a + (b + c')) = S((a + b) + c') \text{ (by the induction hypothesis)} \\ &= (a + b) + S(c') = (a + b) + c. \end{aligned}$$

Therefore, P(c) is true. and by the principle of mathematical induction we have that $a + (b + c) = (a + b) + c$ for all values of $a, b, c \in \mathbb{N}$. □

We previously defined the order on the natural numbers using set inclusions. This order can be rewritten using operations by noting that $m < n$ if, and only if, $n = m + k$ for some natural number k. We will explore how this order interacts with the operations of addition and multiplication.

Lemma 4.1. *Let $m, n, k \in \mathbb{N}$. Then*

$$(1)\ (m = n) \Leftrightarrow (m + k = n + k) \quad \textit{and} \quad (2)\ (m < n) \Leftrightarrow (m + k < n + k).$$

Proof. We will prove part (1) using an induction argument on k using the statement that given $m, n \in \mathbb{N}$, $P(k)$ is the statement that $(m = n) \Leftrightarrow (m + k = n + k)$.

Since 0 is the additive identity, we know that $P(0)$ is true.

Using the definition of $S(n)$, we have that $(m = n) \Leftrightarrow S(n) = S(m)$, showing that $P(1)$ is true.

If we assume for some $l \in \mathbb{N}$ that $P(l)$ is true, then $m+(l+1) = n+(l+1)$ is equivalent to $(m+l)+1 = (n+l)+1$, and since $P(1)$ is true this is equivalent to $(m+l) = (n+l)$. So $P(l)$ being true implies that $P(l+1)$ is true. Therefore, $P(k)$ is true for all $k \in \mathbb{N}$.

In order to prove part (2), we note that

$$\begin{aligned} (m < n) &\Leftrightarrow (\text{There exists } j \in \mathbb{N} \text{ such that } (n = m + j)) \\ &\Leftrightarrow (\text{There exists } j \in \mathbb{N} \text{ such that } ((n + k) = (m + j) + k)) \\ &\Leftrightarrow (\text{There exists } j \in \mathbb{N} \text{ such that } ((n + k) = (m + k) + j)) \\ &\Leftrightarrow (m + k < n + k). \end{aligned}$$

□

Lemma 4.2. *Let $m, n, k \in \mathbb{N}$. Then*

$$(1)\ (m = n) \Leftrightarrow (m \times k = n \times k) \quad \textit{and} \quad (2)\ (m < n) \Leftrightarrow (m \times k < n \times k),\ \textit{for } k > 0.$$

The proof of this lemma is very similar to the previous lemma and is left as an exercise.

4.1.2 Integers

Now that we have defined the set of the natural numbers ($\mathbb{N}$) and the operations on $\mathbb{N}$ of addition and multiplication, we would like to have additive inverses. However, we do not assume that the set of additive inverses exist, but instead construct the integers from the natural numbers.

Subsection 2.2.3 gives the existence of the set

$$\mathbb{N} \times \mathbb{N} = \{(a, b) | a, b \in \mathbb{N}\}.$$

On this set of ordered pairs, we will define an equivalence relation that

$$(a, b) \sim (c, d) \quad \Leftrightarrow \quad a + d = b + c.$$

Since this is an equivalence relation we can define $\mathbb{Z}$ to be the set of equivalence classes under this relation. We will denote by $[(a, b)]$ the equivalence class containing (a, b).

We can now define the operations of addition and multiplication on this set of equivalence classes.

$$[(a, b)] + [(c, d)] := [(a + c, b + d)] \quad \text{and} \quad [(a, b)] \times [(c, d)] := [(ac + bd, ad + bc)]$$

We need to make sure that these definitions are well-defined in that the operations do not depend on which element of the equivalence class is chosen for the representation. Assume that (a, b) and (a', b') are both elements of $[(a, b)]$ and that (c, d) and (c', d') are both elements of $[(c, d)]$. This is equivalent to the statements that

$$a + b' = b + a' \quad \text{and } c + d' = d + c'.$$

So, using properties of equivalent equations involving natural numbers we can add the two equations together and see that

$$(a + b') + (c + d') = (b + a') + (d + c').$$

Using the associative and commutative properties of addition for the natural numbers we see that this equation is equivalent to

$$(a + c) + (b' + d') = (b + d) + (a' + c')$$

and so $[(a + c, b + d)]$ is equivalent to $[(a' + c', b' + d')]$, proving that addition is well-defined. The proof that multiplication is well-defined is left as an exercise.

We can also put an order on this set by defining that

$$[(a, b)] < [(c, d)] \Leftrightarrow a + d < b + c.$$

We can think of the natural numbers, $\mathbb{N}$, as the subset of $\mathbb{N} \times \mathbb{N}$ of the form

$$\mathbb{N} := \mathbb{N} \times \{0\} = \{[(a, 0)] | a \in \mathbb{N}\},$$

since $[(a, 0)] + [(b, 0)] = [(a + b, 0)]$ and $[(a, 0)] \times [(b, 0)] = [(ab, 0)]$.

With this identification of the natural numbers with $\mathbb{N} \times \{0\}$, where $a = [(a, 0)]$, we are able to see that the negative integers can be associated with $\{0\} \times \mathbb{N}$, since $[(a, 0)] + [(0, a)] = [(a, a)] = [(0, 0)]$. We then label $[(0, a)]$ as $-[(a, 0)]$, and usually write it as $-a$. So -4 would be associated to $[(0, 4)]$.

We can then identify every element of $\mathbb{N} \times \mathbb{N}$ with an integer,

$$[(a, b)] \leftrightarrow a + (-b).$$

Thus we have constructed a new set $\mathbb{Z}$ with the operations of addition and multiplication that satisfy the properties for $a, b, c \in \mathbb{Z}$ in Table 4.1.

Table 4.1: Properties of Integers

		Property
$a + b$ is an integer	$a \times b$ is an integer	Closure
$a + (b + c) = (a + b) + c$	$a \times (b \times c) = (a \times b) \times c$	Associative Property
$a + b = b + a$	$a \times b = b \times a$	Commutative Property
$a + 0 = a$	$a \times 1 = a$	Identities
$a \times (b + c) = (a \times b) + (a \times c)$	$(a + b) \times c = (a \times c) + (b \times c)$	Distributive Property
$a + (-a) = 0$		Additive Inverses
If $a, b > 0$, then $ab > 0$.	If $a, b < 0$, then $ab > 0$.	
$ab = 0$ if, and only if, $a = 0$, $b = 0$, or both		No zero divisors

While the integers do not satisfy the Well-Ordering Property (3.4), since the negative numbers are a non-empty set without a least element, it does satisfy the least-upper-bound property that any non-empty set that is bounded above has a least upper bound.

The integers do however, have generalizations of Lemmas 4.1 and 4.2, with the proofs following directly from various properties listed in Table 4.1.

Theorem 4.3. *Let $m, n, k \in \mathbb{Z}$. Then*

$$(1)\ (m = n) \Leftrightarrow (m + k = n + k), \quad (2)\ (m < n) \Leftrightarrow (m + k < n + k),$$

$$(3)\ (m = n) \Leftrightarrow (m \times k = n \times k),\ \textit{with } k \neq 0 \quad \textit{and} \quad (4)\ (m < n) \Leftrightarrow (m \times k < n \times k),\ \textit{with } k > 0.$$

4.1.3 Properties of Exponents

Let a be an integer. We define a^n for each natural number n using the following recursive definition.

$$a^0 := 1, \quad a^1 := a, \quad \text{and} \quad a^{n+1} := a \cdot a^n \ \forall n \in \mathbb{N}.$$

Notice that we have defined $0^0 = 1$. This is not a universally accepted definition, but is widely accepted in the context of algebraic and combinatoric contexts, while left as indeterminate in a limit context of analysis, since

$$\lim_{x \to 0^+} x^0 = 1 \quad \text{and} \quad \lim_{x \to 0^+} 0^x = 0.$$

Related Content Standards

- (6.EE.1)] Write and evaluate numerical expressions involving whole-number exponents.

With this definition, we are able to prove many of the properties of exponents of natural numbers with a base of an integer. It is already given that $a^0 = 1$ and $a^1 = a$.

Lemma 4.3. *Let a be an integer, and let m and n be natural numbers. Then*

$$a^m \cdot a^n = a^{m+n}.$$

Proof. That we are trying to prove something to be true for all natural numbers points us to a proof by induction argument. In this case, we will fix $n \in \mathbb{N}$ and let $P(m)$ be the statement that $a^m \cdot a^n = a^{m+n}$.

Then $P(0)$ is the statement that $a^0 \cdot a^n = a^{0+n}$. Since $a^0 = 1$ and $0 + n = n$, we have that $1 \cdot a^n = a^n$, which is a true statement.

If we assume that $P(k)$ is true for some value of k (we already know that it is true for $k = 0$), then $a^k \cdot a^n = a^{k+n}$. This means that

$$\begin{aligned} a^{(k+1)} \cdot a^n &= (a \cdot a^k) \cdot a^n \text{ (by the definition of the exponent)} \\ &= a \cdot (a^k \cdot a^n) \text{ (since multiplication is associative for integers)} \\ &= a \cdot a^{k+n} \text{ (by the assumption that } P(k) \text{ is true)} \\ &= a^{(k+n+1)} \text{ (by the definition of the exponent)} \\ &= a^{(k+1)+n} \text{ (since addition is associative for the natural numbers).} \end{aligned}$$

So, $P(k+1)$ is true.

Therefore, by induction, $P(m)$ is true for all natural numbers m. □

Lemma 4.4. *Let a and b be any integers, and let n be any natural number. Then*

$$(ab)^n = a^n \cdot b^n.$$

Proof. Since we are proving a statement to be true for all natural numbers, we will use a proof by induction with the statement $P(n)$ being that $(ab)^n = a^n \cdot b^n$.

Since $P(0)$ is the statement that $(ab)^0 = a^0 \cdot b^0$, we have that this is equivalent to the statement that $1 = 1$, and so is true.

If we assume that $P(k)$ is true for some value of k, then we have that $(ab)^k = a^k \cdot b^k$. We can then use properties of addition and multiplication of integers, and the definition of exponents, to see that

$$(ab)^{(k+1)} = (ab)^k \cdot (ab) = a \cdot (ab)^k \cdot b = (a \cdot a^k) \cdot (b^k \cdot b) = a^{(k+1)} \cdot b^{(k+1)}$$

and so $P(k+1)$ is true.

Therefore, by induction, $P(n)$ is true for all $n \in \mathbb{N}$. □

Lemma 4.5. *Let a be an integer and n and m be any natural numbers. Then*

$$(a^m)^n = a^{mn}.$$

The proof of this lemma is similar to the proof of Lemma 4.3 and so will be left as an exercise.

Lemma 4.6. *Let a and b be integers with $0 < a < b$. Then for each natural number $n \geq 1$, $0 < a^n < b^n$.*

Proof. We will use an induction argument for fixed integers a and b, with $a < b$. We let $P(n)$ be the statement, $0 < a^n < b^n$. Since $a^1 = a$ and $b^1 = b$, we see that $P(1)$ is true. If for some $k \in \mathbb{N}$, with $k \geq 1$, $P(k)$ is true, then

$$a^{(k+1)} = a^k \cdot a^1 < b^k \cdot a^1 < b^k \cdot b^1 = b^{(k+1)}$$

using the induction hypothesis and Lemma 4.2. Since the natural numbers are closed under multiplication (the product of two natural numbers is a natural number), and therefore also under exponents, we have that $P(k+1)$ is true. Therefore, $P(n)$ is true for all $n \in \mathbb{N}$, with $n \geq 1$. □

We will summarize these results in the following theorem and extend these properties to other number systems throughout the remainder of the chapter.

Theorem 4.4. *Let a and b be integers.*

a. $a^0 = 1$

b. $a^1 = a$

c. $a^m \cdot a^n = a^{m+n}$ *for each* $m, n \in \mathbb{N}$

d. $(ab)^n = a^n \cdot b^n$ *for each* $n \in \mathbb{N}$

e. $(a^m)^n = a^{mn}$ *for each* $m, n \in \mathbb{N}$

f. *If* $0 < a < b$, $0 < a^n < b^n$ *for each* $n \in \mathbb{N}$

4.1.4 Exercises

1. After introducing the associative property for addition, a student asks if there is something similar for subtraction so that
$$A - (B - C) = (A - B) - C.$$
How would you respond to the student?

2. Prove that the relation on $\mathbb{N} \times \mathbb{N}$ defined by
$$(a, b) \sim (c, d) \quad \Leftrightarrow \quad a + d = b + c$$
is an equivalence relation.

3. Prove Lemma 4.2.

4. Prove Lemma 4.5.

5. Describe the process of proofs by induction in your own words.

4.2 Representations of Integers

Implicit to our discussion of number systems is the idea of representation. While Chapter 3 focused on the concepts of equivalence and equality, representation is concerned with how we choose to communicate mathematical ideas. As advanced students of mathematics, it is easy to take for granted the ways that we communicate fundamental notions such as number or operation using symbols like "3." However, the numbers and symbols we take for granted are choices that have been made and agreed upon.

While the choices often allow for efficient communication, they are often not the most obvious choices for developing conceptual understanding of the abstract objects they represent. For example, when was the last time you saw negative 3 things in the natural world? While we understand the notion of "I owe 3 dollars," the concept of owing money is an abstract concept, not a physical one. It takes a long time for students to develop the conceptual understanding of how to work with mathematical objects they cannot physically combine or otherwise manipulate. Fortunately, because the negative sign is a choice of one among many, we can choose different representations of negative numbers and other abstract mathematical objects to facilitate our teaching. This section examines some of the other choices we can make for representing our numeration system and the four common operations.

4.2.1 Numeration Systems

When we write the number 13, 436, we are communicating how many of something we have. To understand just how many things 13, 436 is, we need two fundamental ideas: digit and place value. For example, in 13, 435 we used the *digit* 3 twice: once in the tens place value and once in the thousands place value. In these cases, the digit 3 communicates how many of each place value we have. Together, the set of digits and a shared understanding of place value gives us our current numeration system.

The Hindu-Arabic Numeration System is the numeration system we use in most of the world today. It is composed of three main ideas:

1. A set of 10 digits: $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$
2. A base 10 place value system
3. Reliance on a digit that means "none"

Note that these are choices: the base of a place value system determines how many digits you need. While our culture has settled on base 10 (based on the number of digits on our hands), the Mayan's developed a base-20 system (based on the number of digits on hands and feet) and computer machine language uses a base 2 system (based on a signal being on or off).

The place value base is used to build the place value by using exponential powers of the base value. For example:

$$1234.567 = 1 \cdot 10^3 + 2 \cdot 10^2 + 3 \cdot 10^1 + 4 \cdot 10^0 + 5 \cdot 10^{-1} + 6 \cdot 10^{-2} + 7 \cdot 10^{-3}.$$

Although this probably all sounds like review and very obvious, these are not obvious choices. It took a long time for human civilizations to develop the base number systems and children have to be indoctrinated into it. It is not an easy idea that "3" in one position does not have the same meaning as a "3" in another position. Although this section focuses mostly on representation of number, the use of symbols taking on multiple meanings in mathematics in different contexts is one with which mathematics teachers must repeatedly grapple. For example, placing two mathematical symbols next to each other could represent either multiplication or addition, i.e. $2x$ and $2\frac{1}{2}$.

To better understand the power of these systems, we will examine a couple numeration systems that have different combinations of the properties of the Hindu-Arabic system.

Example 4.1 (Egyptian Numeration)**.** The Egyptian numeration system uses individual Hieroglyphs to hold value, with new symbols needed for each power of 10). To communicate multiples of a particular power of ten, the Egyptians used repeated Hieroglyphs. For example, the water lily or lotus Hieroglyphs means 1, 000. A sequence of three water lily Hieroglyphs would mean 3, 000. There are no requirements on how Hieroglyphs are presented, so the Hieroglyphs communicating "1" could appear at any place in the expression for the number, and did not even have to appear together. A clear con of this system is that new symbols are needed for each power of 10 would require a new Hieroglyph. In addition, the lack of an ordering or organizational system would make large numbers potentially cumbersome to read.

Value	1	10	100	1,000	10,000	100,000	many
Hieroglyph	𓏺	𓎆	𓍢	𓆼	𓂭	𓆐	𓁨

So the number 3, 244 would be written as

𓆼𓆼𓆼 𓍢𓍢𓎆𓎆𓎆𓎆 𓏺𓏺𓏺𓏺

Example 4.2 (Roman Numeration)**.** The Roman numeration system is similar to the Egyptian system, in that the Roman numbers are each unique to a particular value (for example, X means 10). The Roman system improved on the Egyptian numeration system by imposing an order to how these numbers could be

Table 4.2: Roman Numeration

Arabic Numeral	Roman Numeral
1	I
10	X
100	C
1,000	M

presented and partially solved the potential for a large number of repeated numbers by adding the subtraction element (so, IX would mean take 1 away from 10, while XI means add one to 10). However, as students of the Super Bowl know, reading these numbers still takes a fair amount of decoding and there is a still a problem with needing a new number for each new grouping.

So the Roman numeral version of $3,244$ would be MMMCCXLIV.

Example 4.3 (Base 5)**.** Base 5 is like our current standard numeration system, except it uses a set of five digits, $\{0, 1, 2, 3, 4\}$, and place values are based on powers of 5, not 10. In the United States we use base 5 in some of our coins, in that there are 5 pennies in a nickel and 5 nickels in a quarter. Then the number 14_5 (read as fourteen base five) is saying 1 nickel and 4 pennies, or 9 cents. Similarly,

$$231_5 = 2 \cdot 5^2 + 3 \cdot 5^1 + 1 \cdot 5^0 = 66_{10}.$$

Our number of $3,244 = 1 \times 5^5 + 4 \times 5^2 + 3 \times 5^1 + 4$ would be 100434_5 in Base 5.

Note: while there is no obvious use for students to learn operations in a different base, working outside of our base 10 system can help teachers recognize and understand the inner workings of the base 10 system and its associated algorithms. As these algorithms are introduced throughout the section, you are encouraged to use the algorithms to perform operations on numbers in these additional base systems.

Example 4.4 (Base 12)**.** The duodecimal (base 12) number system is a component of some languages in West Africa and Nepal and is used in our everyday measurements of dozen and gross. A duodecimal system requires twelve digits, $0, 1, 2, 3, 4, 5, 6, 7, 8, 9, t, e$, and place values are based on powers of 12. In this system $6 + 5 = e$, $7 + 8 = 13$, and

$$3t \times e4 = 2te4.$$

So

$$3,244 = 1 \times 12^3 + 10 \times 12^2 + 6 \times 12 + 4 = 1t64_{12}.$$

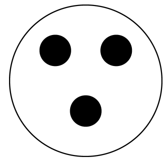
One of the implications of a base 12 system is additional tricks for multiplying numbers due to a larger number of factors of twelve than ten.

4.2.2 Representations of Number

In talking about numeration systems, we have exclusively been using one form of number representation. However, numeration systems are only one way that we communicate ideas about number. In this section, we explore different ways of representing numbers, each which helps develop different properties of numbers. For this introduction to representation, we focus first on the positive natural numbers, adding additional elements to these representations as we develop the various number systems.

Sets

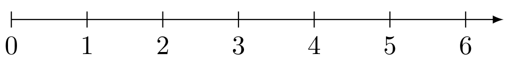
Sets can be used to represent numbers, either pictorially as a circle containing generic objects or as a set with general elements, such as the figure below or $\{a, b, c\}$, respectively.



Using sets offers the most general representation of number we can develop, as it does not impose any structure on the representation beyond the property of “how many”. Relationships between numbers (such as inequality or equality) are fairly straightforward to model, but comparing a large number at one time would be difficult using this representation. As a result, most middle and high school textbooks do not represent numbers this way, although it is fairly common in the elementary grades.

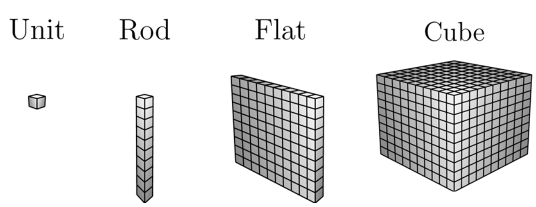
Number Line

The number line imposes more structure on the number representation than the use of sets. In particular, it organizes numbers along a line, with integers spaced evenly along the line. This representation allows multiple numbers to be visualized at the same time while also showing relationships between the numbers represented. In this representation, any number that falls at the location of 3 is equivalent to three. Numbers of the left of 3 are less than it, while numbers above are greater.



Base-10 Blocks

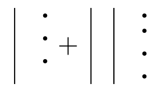
Base-10 blocks are usually a physical way (although virtual and pictorial versions exist) to represent numbers specifically designed to promote understanding of place value. The blocks come in four sizes: units (small, approximately 1-centimeter cubes), rods (10, represented by a stack of 10 unit cubes), flats (100, represented by a 10 by 10 grid of unit cubes), and big cubes (1000, represented by a cube with side lengths equal to 10 unit cubes). Representing a number using base-10 blocks means drawing or collecting the correct number of blocks for each place value.



In order to write down the reasoning these blocks are sometimes transitioned into dots, lines, and squares. So that $13 + 24$ is represented by

Cuisenaire Rods.

Cuisenaire rods are another physical way that numbers are represented. They are a collection of long prisms of different colors and lengths. Rods of the same length are the same color, and each length is a natural number multiple of the smallest rod. These rods are used for a variety of reasons including factoring of natural numbers, but one of the most common is to help students understand concepts related to rational numbers and the operation of division.



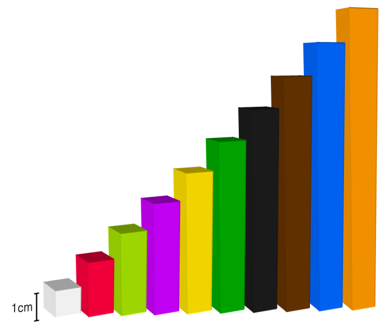
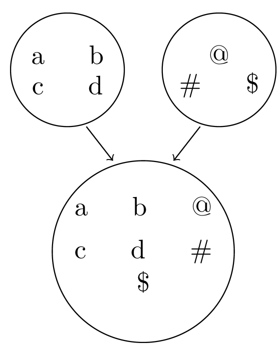


Figure 4.1: Cuisenaire Rods

4.2.3 Addition Models and Algorithms

Much like the concept of number, it is easy to overlook how quickly addition becomes a complex endeavor for young students of mathematics. Fundamentally, addition represents the notion of "combine' ' or finding the union of mutually disjoint sets.

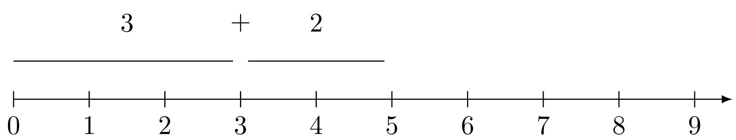


Typically, we quickly move away from representing addition using sets and instead use the number line. For natural number addition, this process is fairly straight forward: Students start by identifying the first addend. Then, count up the number that is being added. The number you land on is the result.

Thinking of addition as the union of sets or on a number line of the natural numbers are good, conceptual starting places for understanding addition. However, doing so will lead some students to think that "when two numbers are added, the result will be the same size or bigger." In the natural numbers this thinking will work, but in the integers the reasoning begins to fall apart. Note that idea of "combine" does not say anything about the size of the result relative to the others, so the common error is a result of students focusing on the result rather than the concept of addition.

When asked to add something like $125 + 258$ without the use of a calculator, most adults are well-trained to rewrite the sum and apply the standard algorithm.

While the work of teaching children these algorithms is usually done in elementary school, it is worth taking a little time to appreciate the ways these traditional methods take advantage of the properties of natural numbers to allow for quick and efficient calculation. To facilitate this discussion, we present several alternative algorithms. It is not our goal to instruct our readers in each of these algorithms: there is not a way to present all possible ways an addition problem can be solved. Rather, we choose algorithms that maximize the opportunity to explore the hidden work of the standard algorithm. A mathematics teacher who studies these examples in relation to the standard algorithm should be able to teach themselves unfamiliar



$$
\begin{array}{cccc}
 & & 1 & \\
 & 1 & 2 & 5 \\
+ & 2 & 5 & 8 \\
\hline
 & 3 & 8 & 3
\end{array}
$$

algorithms that they encounter in the wild. For each of these, we use the example $125 + 258$.

Expanded form

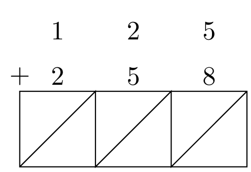
We can rewrite our numbers using place value as $(100 + 20 + 5)$ and $(200 + 50 + 8)$ and then use associative and commutative properties of addition for the natural numbers.

$$
\begin{aligned}
125 + 258 &= (100 + 20 + 5) + (200 + 50 + 8) \\
&= (100 + 200) + (20 + 50) + (5 + 8) \\
&= (100 + 200) + (20 + 50) + (13) \\
&= (100 + 200) + (20 + 50) + (10 + 3) \\
&= (100 + 200) + (20 + 50 + 10) + (3) \\
&= 300 + 80 + 3 = 383
\end{aligned}
$$

Lattice Algorithm

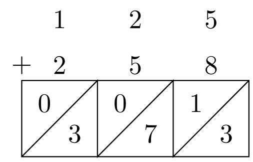
The lattice algorithm described below provides another method of adding two numbers by keeping track of place value through some implicit methods.

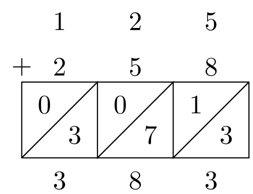
- Stack the two numbers so that corresponding place values are in the same column. Draw the traditional equals bar.



- Starting with the right-most column, add within each column. The result of the digit addition will be a number between 0 and 19. The resulting number is written into the square below the column, with the higher place value going into the higher triangle.

- Add down each diagonal. You are in effect adding the total number of items in each place value.





Expanded algorithm

The expanded algorithm makes the implicit structure of the lattice algorithm more explicit.

$$\begin{array}{r} 125 \\ +258 \\ \hline 13 \\ 70 \\ +\ 300 \\ \hline 383 \end{array}$$

You may have noticed that regardless of which method we used, there were several common elements that all of the algorithms created in one form or another. Perhaps most obviously, 13 occurred in all three calculations. The intermediate sums an algorithm produces are called **partial sums**. Note that although every algorithm used partial sums, they are not always the same between algorithms, even for the same addition problem. These differences in partial sums are a result of where the algorithm does the **regrouping** of the partial sums. Regrouping, generally, is the process of converting numbers expressed in terms of one place value into a different place value. For example, in the standard algorithm example at the beginning of this section, we got 13 when we added the ones digits in the right most column. Since 13 ones is not useful for continued adding, we regrouped 13 as 1 ten and 3 ones. We wrote the 3 in the ones place value and the 1 above the 10's place value column. You may have learned this process by the name of "carrying the one". This language is no longer encouraged in teaching as it masks the fact that the value of the one changes as a person works through every place value.

4.2.4 Subtraction Models and Algorithms

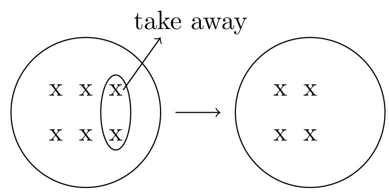
Conceptually, we described addition as the process of combining. Subtraction has two different physical interpretations: the process of removal (often referred to as taking away) or the inverse operation of addition. These are clearly related, but are developed separately, which we explore a bit more in the following paragraphs. Although both interpretations of subtraction seem straight forward when modeled in the physical world, students struggle with this operation a great deal more than addition. This may be at least partially because while the natural numbers have some great properties under addition (closed, commutative, and associative) that make addition easy to set up, but none of these properties under subtraction.

Using our different representations, we can model subtraction a couple different ways: the take away model or the missing addend model.

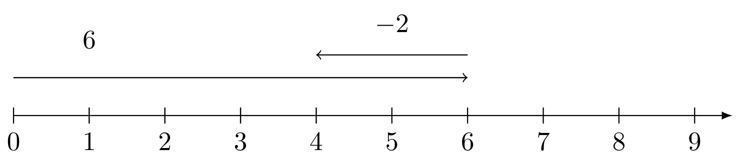
Take-Away Model

The first interpretation of subtraction is as removal or the process of taking away. In the mathematics education literature, this is called the "take-away model". This conceptualization of subtraction is a direct interpretation of the calculation $a - b$, where the translation of this expression is "a take away b."

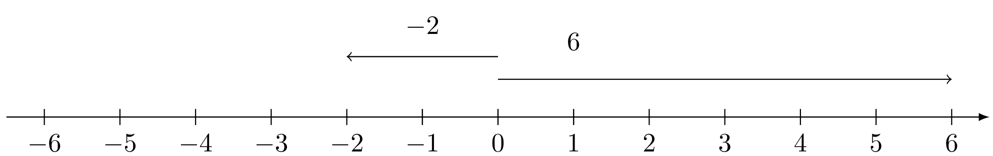
To model $a - b$ using sets and the take-away model, we start with a set containing a elements and remove b elements from it. The composition of the set after the removal process is the resulting difference.



To model $a - b$ using the number line and the take-away model, we find a on the number line and then move back b units. The place where you land is the resulting difference.



Once we have introduced the integer number system, we extend the number line and think of $a-b$ as $a+(-b)$ where $-b$ is viewed as a vector of length $|b|$ pointed in the negative direction.



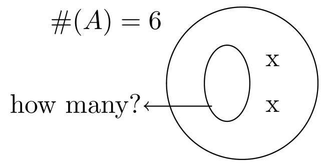
Related Content Standards

- (6.NS.6) Understand a rational number as a point on the number line. Extend number line diagrams and coordinate axes familiar from previous grades to represent points on the line and in the plane with negative number coordinates.
    a. Recognize opposite signs of numbers as indicating locations on opposite sides of 0 on the number line; recognize that the opposite of the opposite of a number is the number itself, e.g., $-(-3) = 3$, and that 0 is its own opposite.

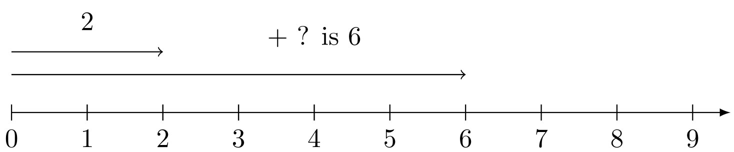
Missing Addend Approach

The missing addend approach interpretation of subtraction is an explicit model of subtraction as the inverse operation of addition, while converting the computationally more nuanced subtraction problem into an addition problem where the number facts and such are easier. In particular, in the missing addend approach, the subtraction expression $a - b$ is assumed to equal some unknown (so $a - b =?$), which we can rewrite as $b+? = a$.

Using the set model and the missing addend approach, we take the statement $b+? = a$ to mean that we know that we want a set to contain a elements and we have b. We must determine how many additional elements we need in order to bring the total number of things in our set up to a.



Using the number line, we reconceptualize $a - b = ?$ As $b + ? = a$ which means we start at b on the number line and then find a. We then ask, how much do I have to move (and in what direction) from b to get to a?. The answer to this question is the answer to the original subtraction problem. This process when formalized into an algorithm that counts up by place value is called **counting up**.



Note that in none of the conceptual work of this section have we used the language of larger number and smaller number, instead focusing on the expression of the difference, $a - b$. Early instruction in subtraction only presents cases of subtraction problems $a - b$ when $a \geq b$ and both are natural numbers. As a result of this presentation bias, students sometimes develop the impression that the results of subtraction will always be smaller than the original number. They may also believe that subtraction when $a < b$ cannot be done. The conceptual work we have developed in this section for the models of subtraction works for most sets of numbers students encounter in the K-12 mathematics curriculum.

As with addition, we have a standard subtraction algorithm. For example, consider the subtraction problem $423 - 368$. A well-trained student asked to do this problem without a calculator will produce work that looks something like the following:

$$\begin{array}{rccc} & 3 & 11 & \\ & \cancel{4} & \cancel{2} & {}^{1}3 \\ - & 3 & 6 & 8 \\ \hline & & 5 & 5 \end{array}$$

As with addition, the traditional subtraction algorithm uses regrouping strategies and the properties of place value to allow for quick calculation in a small amount of space. (You may have learned the regrouping strategy under the name "borrowing" but, as with addition, this name is no longer used for similar reasons.)

We next provide several alternative algorithms for subtraction designed to help develop conceptual understanding of the subtraction algorithm.

Equal Addition

The equal addition algorithm repeatedly adds zero to the problem in a way that makes the subtraction easier to compute. The goal is to write an equivalent problem that has at least one zero place value.

$$\begin{aligned} 423 - 368 &= 423 - 368 + (2 - 2) = 425 - 370 \\ &= (425 - 370) + (30 - 30) = 455 - 400 \\ &= 55 \end{aligned}$$

Partial Differences

Another method of subtraction is to subtract from left to right with partial differences. In this case, we subtract the smaller digit from the larger digit, keeping track of positive or negative.

$$
\begin{array}{cccc}
 & 4 & 2 & 3 \\
- & 3 & 6 & 8 \\
\hline
+ & 1 & 0 & 0 \\
 & & & \\
 & & & \\
\end{array}
\qquad
\begin{array}{cccc}
 & 4 & 2 & 3 \\
- & 3 & 6 & 8 \\
\hline
+ & 1 & 0 & 0 \\
- & & 4 & 0 \\
 & & & \\
\end{array}
\qquad
\begin{array}{cccc}
 & 4 & 2 & 3 \\
- & 3 & 6 & 8 \\
\hline
+ & 1 & 0 & 0 \\
- & & 4 & 0 \\
- & & & 5 \\
\end{array}
\qquad
\begin{array}{cccc}
 & 4 & 2 & 3 \\
- & 3 & 6 & 8 \\
\hline
+ & 1 & 0 & 0 \\
- & & 4 & 0 \\
- & & & 5 \\
\hline
 & & 5 & 5 \\
\end{array}
$$

In the section on subtraction, we commented that when students first learn subtraction in the natural numbers, they are usually only given examples of the form $a - b$ where $b \leq a$. Part of the reason for only having students work on these types of problems is that children are not born knowing about negative numbers. The only place we "see' ' negative numbers in the real world is when humans have imposed a scale on something, such as temperature. However, this is not a true physical representation, since we have simply imposed a scale on a naturally occurring thing, where for one reason or another, we want to have a zero reference point.

To allow us to find answers to subtraction problems where $a < b$, we expand our number system to the integers. You may have noticed that we use "$-$" in the above paragraph to mean additive inverse of a positive number. In the paragraph before that we used the same symbol to mean subtraction. In reality, these are fairly different concepts. Subtraction is an operation. Additive inverse is a relation. Long-time students of mathematics know that these interpretations are related and will move between them fairly seamlessly, but they are not the same thing.

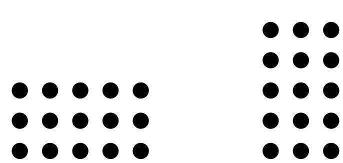
4.2.5 Multiplication Models and Algorithms

Repeated Addition

Most commonly, multiplication is described as **repeated addition**. For example, the product 2×4 can be understood as adding 2 four times $(2 + 2 + 2 + 2)$, while the product 4×2 can be understood as adding 4 two times $(4 + 4)$. While conceptually, this link between addition and multiplication is important, it quickly loses efficiency in practical applications (e.g., it would be cumbersome but possible to multiply 23×345 using repeated addition). That said, it is not uncommon to see secondary students apply this approach to multiplication problems.

Rectangular Array

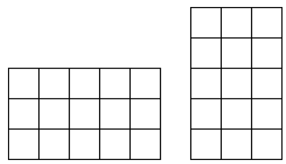
Another way to represent multiplication is through a **rectangular array**. One way to think about this is counting the number of chairs in a rows and b columns. Such arrays are discrete sets. We can see below how one can use two arrays to represent 3×5 and 5×3.



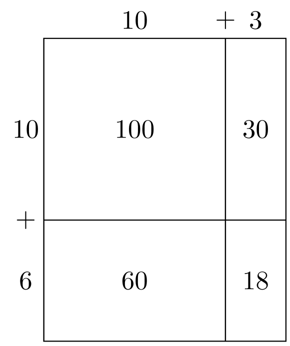
We can see that the two arrays are related to each other by a rotation. In other words, the order of multiplication is simply a different perspective on the same array, giving the commutative property of multiplication.

Area Model

While arrays work well for representing multiplication of natural numbers, they do not work well for integers, rational numbers, or real numbers. So we generalize the array model to an **area model**. If we change the dots in the array to blocks of 1 square unit measure, the above arrays become rectangles with square tiles helping students see the connection between the area of a rectangle and multiplication.



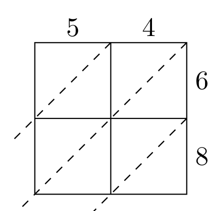
This area model can then be generalized to help students multiply multi-digit natural numbers and the relationship to the distribution property.



$13 \times 16 = 208$

Lattice Multiplication

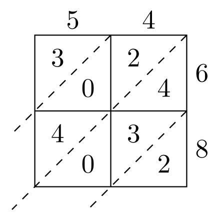
The area model can become unwieldy with larger numbers. The lattice multiplication algorithm is a more abstract view of the area model by using a lattice to keep track of place values. Here we can see the progression of using a lattice model to multiply $54 \times 68 = 3672$.

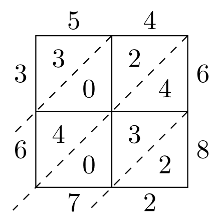


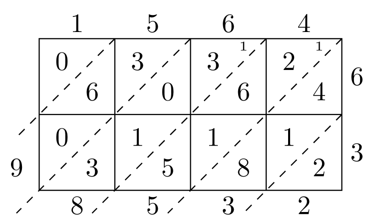
After setting up the lattice with diagonals, we multiply all of the partial products and enter the product in the corresponding square with the one's digit in the bottom right and the ten's digit in the upper left. After all of the partial products are complete, we add along the diagonals in order to find the number of units of that place value. For some products, the sum along the diagonal is larger than 9 and so we have to regroup as is the case for 1564×63.

Partial Products

We continue the progression of creating more abstract algorithms for multiplication by removing some of the explicit structure using the **partial products algorithm**.







$$
\begin{array}{cccccc}
 & & 1 & 5 & 6 & 4 \\
 & \times & & & 6 & 3 \\
\hline
 & & {\scriptscriptstyle 1} & {\scriptscriptstyle 1} & 1 & 2 \\
 & & & 1 & 8 & 0 \\
 & & 1 & 5 & 0 & 0 \\
 & & 3 & 0 & 0 & 0 \\
 & & & 2 & 4 & 0 \\
 & & 3 & 6 & 0 & 0 \\
 & 3 & 0 & 0 & 0 & 0 \\
+ & 6 & 0 & 0 & 0 & 0 \\
\hline
 & 9 & 8 & 5 & 3 & 2
\end{array}
$$

Standard U.S. Multiplication Algorithm

If we continue down the path of more abstract algorithms we come to what is considered the standard algorithm for multiplication. In this algorithm we compress the information in the partial products algorithm into many fewer lines.

$$\begin{array}{ccccc} & 1 & 5 & 6 & 4 \\ \times & & & 6 & 3 \\ \hline & 4 & 6 & 9 & 2 \\ 9 & 3 & 8 & 4 & \\ \hline 9 & 8 & 5 & 3 & 2 \end{array}$$

However, with this reduction in the number of lines comes an increase in the number of things that one must keep track of implicitly, often resulting in an increase in mistakes or a lack of understanding underlying concepts.

4.2.6 Division Models and Algorithms

Division is the fourth major operation. Conceptually, the division problem $a \div b$ can be viewed several ways: as repeated subtraction, as the inverse of multiplication (though the integers do not contain multiplicative inverses), and as a partition. Regardless of the model, all three of these representations draw on the notion that when we divide a by b ($a \div b$) the division operation produces two unique numbers: the **quotient**, q, and the **remainder**, r, where $0 \leq r < b$. (We will go into this division algorithm more in-depth in Section 7.1.) We call b the **divisor** and a the **dividend**. These four quantities are related by $a = bq + r$.

Repeated Subtraction Model

In the repeated subtraction model of division, the division problem $a \div b$ conceptually is thought of as repeatedly subtracting b from a until you no longer have a number bigger than b. The quotient is the number of times you could subtract and the remainder is what is left over after the quotient has been determined.

Missing Factor Model

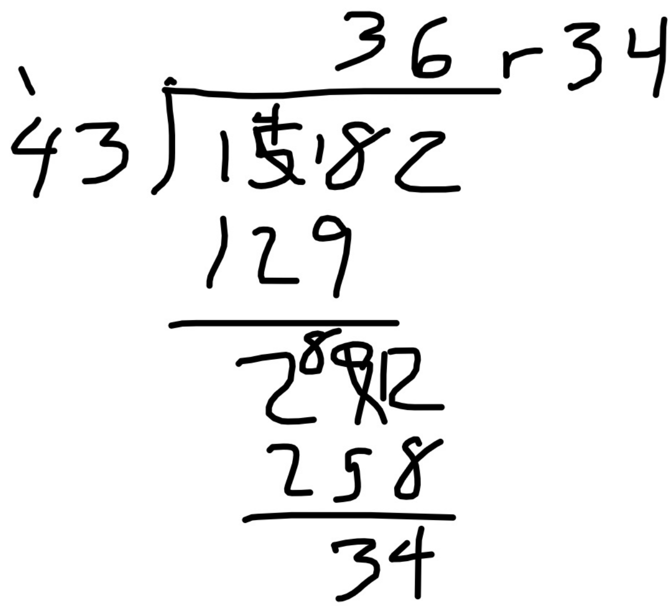
In the missing factor model, the division problem defines division as the inverse operation of multiplication. The quotient and remainder are not separated. Instead, we reorganize the quantities in their reciprocal relationships in the equation $a = bq + r$. Generally, we use the missing factor approach when $r = 0$. In particular, if $a \div b = q$, where q is a unique value, then we rewrite the problem as the product $bq = a$.

Set (Partition) Model

The set (partition) model frames the quotient as a partition of the whole, a, into b even parts. In particular, $a \div b$ is thought of as asking if I must evenly distribute a between b entities, how much will each entity get? While in practice, the finding of the quotient for both the partition and repeated subtraction models are similar, conceptually, the two methods handle the remainder differently. The repeated subtraction approach reports the remainder, while the partition model invites the person completing the problem to keep going until all parts of a, including r, have been evenly distributed.

4.2.6.1 Division Algorithms

The long division algorithm is the last major algorithm students learn. If we examine the work of the long division algorithm, presented below, a few things are apparent. First, the algorithm is the only one that does not begin work with the ones place value, making it an anomaly in students' learning. Secondly, it relies on students to have mastered the subtraction algorithm. And lastly, when we have multi-digit divisors, estimation is an important prerequisite skill. These factors all make the division algorithm particularly tricky to master.



With the widespread availability of cheap and accurate calculators, it is worth asking why we still teach the division algorithm (or really, most advanced arithmetic). And indeed, long division is not emphasized in the curriculum as much as it used to be for this very reason. We do provide a brief treatment here as the division algorithm, while not widely used in daily life, does lay a foundation for polynomial division, discussed in later chapters. In partial products algorithm presented below, pay particular attention to how the various division models are codified into the structure of the algorithm and how the partial quotients differ or are similar to those of the standard algorithm.

43	1	5	8	2	
	−	4	3	0	10
	1	1	5	2	
	−	4	3	0	10
		7	2	2	
	−	4	3	0	10
		2	9	2	
	−	2	1	5	5
			7	7	
		−	4	3	1
			3	4	36

4.2.7 Exercises

1. Use the definition of base 5 in this section to answer the following questions.

   a. What is the base 5 number following 4_5?
   b. Express 234_5 in base 10.
   c. Express 234_{10} in base 5.

2. Add $342_5 + 134_5$ using each of the four addition algorithms (expanded form, lattice algorithm, expanded algorithm, and standard algorithm) presented in this section.

3. Complete the product $234_5 \times 32_5$ using each of the three multiplication algorithms presented in this section (lattice, partial products, standard).

4. Complete the division problem $341_5 \div 41_5$ using any division algorithm discussed above.

5. Construct the addition and multiplication tables for the first twelve digits in a base 12 system. What patterns do you notice?

6. Read up on the Mayan numeration system, which is a base 20 system.

   a. Describe the patterns used to develop the 20 digits.
   b. How is this system comparable to the numeration systems discussed in this section.

7. For the two-number lattice addition, will adding along a diagonal ever require a student to regroup? Why or why not?

8. The standard addition algorithm requires that a person adds columns right to left. Many students just learning the algorithm will make the mistake of starting with the left most column and working towards right.

   a. What conceptual understanding is a student missing when they make this type of error?
   b. Which of the three non-standard algorithms will result in a correct sum regardless of whether a person adds right to left or left to right?
   c. Which of the four algorithms presented in this section might help students develop the conceptual understanding you identified in part (a)?

9. Another common error students make with the standard addition algorithm is that students will line up the first digit of a number with the first digit of the second number, regardless of whether or not the numbers have the same number of digits.

   a. What conceptual understanding is a student missing when they make this type of error?
   b. Which of the four algorithms presented in this section might help students develop the conceptual understanding you identified in part (a)?

10. Using a set of base-ten blocks (or a virtual base-ten block simulation), model the addition problem $348 + 753$.

    a. Write out the arithmetic steps you went through in completing the addition using the blocks.
    b. Describe how base ten blocks might support conceptual understanding of the standard addition algorithm.

11. How does the standard multiplication algorithm utilize each of the four properties of the natural numbers (commutative, associative, distributive, identity)?

12. Why, in multiplication by multi-digit numbers in the standard multiplication algorithm, do you sometimes include zeros in the right most digit places?

13. For each model of division (e.g., repeated subtraction, partitions, missing factor), explain why the quotient $a \div 0$ does not make sense.

4.3 Number Theory

Now that we have defined the natural numbers and the integers, we will study some of the properties of these number systems. Some of the properties of the integers, including the Fundamental Theorem of Arithmetic, we will study more in Section 7.1.

One of the core properties of the integers is the division algorithm that states that for a pair of integers a and b, with $b \neq 0$, there is a unique quotient and remainder associated to them.

Theorem 4.5 (The Division Algorithm for Integers)**.** *If a and b are integers with $b > 0$, then there are unique integers q and r such that*

$$a = bq + r, \quad \textit{with } 0 \leq r < b.$$

Proof. Let a and b be integers with $b > 0$. We will first prove the existence of integers q and r such that

$$a = bq + r, \quad \text{with } 0 \leq r < b$$

and then prove that these integers are unique.

- **Existence.** If $a \geq 0$, then we can let

  $$S = \{a - kb \mid a - kb \geq 0 \text{ and } k \text{ is a non-negative integer}\}.$$

  Since $a \geq 0$, we see that $a = a - 0 \cdot b$ is an element of S and so S is non-empty. We also see that $S \subseteq [0, a]$.

  So by the Well-Ordering Property of the Integers, there is a smallest element of S which we will call r. Since $r \in S$, there exists a non-negative integer q such that $r = a - qb$.

  Since r is the smallest element of S and since $a - (q+1)b < a - qb = r$, we see that $a - (q+1)b < 0$, otherwise it would be a smaller element of S.

  This implies that $r = a - qb < b$. Therefore we have the existence of the q and r.

  If $a < 0$, using the above process we can find non-negative integers q and r, with $0 \leq r < b$, such that $-a = bq + r$. So $a = b(-q) + r$ satisfies the existence statement.

- **Uniqueness.** In order to prove the uniqueness of the quotient and remainder, we assume that there are two pairs of integers (q_1, r_1) and (q_2, r_2) such that

  $$a = bq_1 + r_1 \text{ and } a = bq_2 + r_2, \text{ with } 0 \leq r_1 < b \text{ and } 0 \leq r_2 < b.$$

  This implies that $bq_1 + r_1 = bq_2 + r_2$ and by rearranging the terms we have that

  $$b(q_1 - q_2) = r_2 - r_1.$$

  If $0 \leq r_1 \leq r_2 < b$, then $0 \leq b(q_1 - q_2) = r_2 - r_1 < b$. So $q_1 - q_2$ is an integer in $[0, 1)$ and so must be 0. Therefore, $q_1 = q_2$ and so $r_1 = r_2$. Therefore, the quotient and remainder are unique.

□

4.3.1 Divisibility

With the division algorithm, for each natural number $n > 1$ we can partition the integers into equivalence classes based on the remainder when divided by n. For example, for $n = 2$ we see that any integer belongs to one of the following sets

$$\{m \in \mathbb{Z} \mid m = 2k \text{ for some } k \in \mathbb{Z}\} \quad \text{or} \quad \{m \in \mathbb{Z} \mid m = 2k + 1 \text{ for some } k \in \mathbb{Z}\},$$

the even and odd integers.

In the following, we will focus on the set

$$\{m \in \mathbb{Z} \mid m = nk \text{ for some } k \in \mathbb{Z}\}.$$

Definition 4.1. Let a and b be integers with $b \neq 0$. We say that that a **is divisible by** b, or a **is a multiple of** b, if there exists a unique $q \in \mathbb{Z}$ such that $a = bq$.

This divisibility has a nice property based on the distributive property of the integers.

Theorem 4.6. *Let a and b be integers that are each divisible by an integer $m \neq 0$. Then $a + b$ is divisible by m.*

Proof. If a and b are divisible by an integer m, we have there there are integer j and k such that $a = mj$ and $b = mk$. So $a + b = mj + mk = m(j + k)$ and so $a + b$ is divisible by m. $\square$

We also see that if a is divisible by b, then $-a$ is also divisible by b.

As students develop fluency with multiplication of integers they often use certain tests to determine divisibility by certain numbers. The most common are given below.

- **Divisibility by 2.** An integer is divisible by 2 if and only if the last digit is 0, 2, 4, 6, or 8.
- **Divisibility by 3.** An integer is divisible by 3 if and only if the sum of the digits of the integer is divisible by 3.
- **Divisibility by 5.** An integer is divisible by 5 if and only if the last digit is 0 or 5.
- **Divisibility by 9.** An integer is divisible by 9 if and only if the sum of the digits of the integer is divisible by 9.
- **Divisibility by 10.** An integer is divisible by 10 if and only if the last digit is 0.
- **Divisibility by 11.** An integer is divisible by 11 if and only if the alternating sum of the digits is divisible by 11.

We notice that all of these divisibility rules are based on the digits of the integer. So in order to understand why these rules work we need a way to denote an integer by its digits. Since we are using a base 10 notation of the integers, we can see that each integer has a unique representation of for form

$$d_0 \cdot 10^0 + d_1 \cdot 10^1 + d_2 \cdot 10^2 + \ldots + d_n \cdot 10^n,$$

with each of the $d_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and the integer usually written with the digits in the reverse order. For example,

$$235,609 = 9 \cdot 10^0 + 0 \cdot 10^1 + 6 \cdot 10^2 + 5 \cdot 10^3 + 3 \cdot 10^4 + 2 \cdot 10^5.$$

So in order to understand the divisibility by 5, we can see that a number n can be written as

$$\begin{aligned} n &= d_0 \cdot 10^0 + d_1 \cdot 10^1 + d_2 \cdot 10^2 + \cdots + d_n \cdot 10^n \\ &= d_0 + 10 \cdot \left(d_1 \cdot 10^0 + d_2 \cdot 10^1 + \cdots + d_n \cdot 10^{n-1}\right) \\ &= d_0 + 5 \cdot \left(2 \cdot \left(d_1 \cdot 10^0 + d_2 \cdot 10^1 + \cdots + d_n \cdot 10^{n-1}\right)\right) \end{aligned}$$

and so n is divisible by 5 if and only if d_0 is divisible by 5. And we have our result because the only digits divisible by 5 are 0 and 5.

Understanding the reasoning behind the rules of divisibility by 3 and 9 are a little different. Each of these rules are based on

$$10^n - 1 = 9 + 9 \cdot 10^1 + \cdots + 9 \cdot 10^{n-1} = 9 \cdot \left(1 + 1 \cdot 10^1 + \cdots + 1 \cdot 10^{n-1}\right),$$

and so $10^n - 1$ is divisible by 9\$. The first few of these are $10 - 1 = 9$, $10^2 - 1 = 99 = 9 \cdot 11$, $10^3 - 1 = 999 = 9 \cdot 111$. So the number n can be written as

$$\begin{aligned}
n &= d_0 \cdot 10^0 + d_1 \cdot 10^1 + d_2 \cdot 10^2 + \cdots + d_n \cdot 10^n \\
&= d_0 + d_1 \cdot (10^1 - 1 + 1) + d_2 \cdot (10^2 - 1 + 1) + \cdots + d_n \cdot (10^n - 1 + 1) \\
&= d_0 + d_1 \cdot (10^1 - 1) + d_1 + d_2 \cdot (10^2 - 1) + d_2 + \cdots + d_n \cdot (10^n - 1) + d_n \\
&= (d_0 + d_1 + \cdots + d_n) + \left(d_1 \cdot (10^1 - 1) + d_2 \cdot (10^2 - 1) + \cdots + d_n \cdot (10^n - 1)\right).
\end{aligned}$$

Since each of the $d_i \cdot (10^i - 1)$ are divisible by 9, and hence divisible by 3, we see that n is divisible by 9 or 3 if and only if $(d_0 + d_1 + \cdots + d_n)$ is divisible by 9 or 3.

For each of these divisibility rules, we see that the relationship of the integer to the base of 10 is the key component. This relationship is explored further in the exercises by studying similar rules for a base 12 system.

4.3.2 Pascal's Triangle

Pacal's Triangle is a triangular array where each entry is the sum of the two entries above that location, see Figure 4.2.

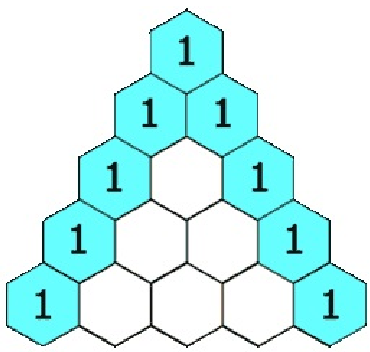


Figure 4.2: Animated Figure of Pascal's Triangle

$$\begin{array}{lccccccccccccccc}
n=0: & & & & & & & & 1 & & & & & & & \\
n=1: & & & & & & & 1 & & 1 & & & & & & \\
n=2: & & & & & & 1 & & 2 & & 1 & & & & & \\
n=3: & & & & & 1 & & 3 & & 3 & & 1 & & & & \\
n=4: & & & & 1 & & 4 & & 6 & & 4 & & 1 & & & \\
n=5: & & & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & & \\
n=6: & & 1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1 & \\
n=7: & 1 & & 7 & & 21 & & 35 & & 35 & & 21 & & 7 & & 1
\end{array}$$

For each $n \in \mathbb{N}$, we define

$$\binom{n}{0} = 1 \quad \text{and} \quad \binom{n}{n} = 1$$

and for $n > 1$ and $1 \le k \le n-1$ we define

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

So $\binom{n}{k}$ is the entry in the nth row and kth column in Pascal's Triangle that we can see below.

$$
\begin{array}{lccccccccccccccc}
n=0: & & & & & & & & \binom{0}{0} & & & & & & & \\
n=1: & & & & & & & \binom{1}{0} & & \binom{1}{1} & & & & & & \\
n=2: & & & & & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & & & & \\
n=3: & & & & & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & & & & \\
n=4: & & & & \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} & & & \\
n=5: & & & \binom{5}{0} & & \binom{5}{1} & & \binom{5}{2} & & \binom{5}{3} & & \binom{5}{4} & & \binom{5}{5} & & \\
n=6: & & \binom{6}{0} & & \binom{6}{1} & & \binom{6}{2} & & \binom{6}{3} & & \binom{6}{4} & & \binom{6}{5} & & \binom{6}{6} & \\
n=7: & \binom{7}{0} & & \binom{7}{1} & & \binom{7}{2} & & \binom{7}{3} & & \binom{7}{4} & & \binom{7}{5} & & \binom{7}{6} & & \binom{7}{7}
\end{array}
$$

Definition 4.2. For $n \in \mathbb{N}$ we define $0! = 1$ and for $n > 0$,

$$n! = n \cdot (n-1) \cdot (n-2) \cdots (2) \cdot (1)$$

Theorem 4.7. *For $n, k \in \mathbb{N}$ with $0 \leq k \leq n$ we have that*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Proof. As with many proofs of statements that involve the natural numbers, we will prove this using the principle of mathematical induction. For each pair of natural numbers n, k with $0 \leq k \leq n$ we let $P(n,k)$ be the statement that

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

For $n = 0$, we know that k must equal 0, since that is the only number for which $0 \leq k \leq 0$. So to prove the $P(0,0)$ is true we see that

$$\frac{0!}{0! \cdot 0!} = \frac{1}{1} = 1$$

and $\binom{0}{0} = 1$.

We can now look at the induction step by assuming that for a natural number $m > 0$ the statement $P(m,k)$ is true for all $0 \leq k \leq m$.

Then for $n = m+1$ and $k = 0$, we see that $\binom{n}{k} = \binom{m+1}{0} = 1$ by definition. Similarly, for $k = m+1$ $\binom{n}{k} = \binom{m+1}{m+1} = 1$, by definition. For $0 < k < m+1$ we see that $\binom{n}{k}$ is defined by

$$\binom{n}{k} = \binom{m+1}{k} = \binom{m}{k-1} + \binom{m}{k}.$$

And by the induction hypothesis we see that

$$\binom{n}{k} = \binom{m}{k-1} + \binom{m}{k} = \frac{m!}{(k-1)!(m-(k-1))!} + \frac{m!}{k!(m-k)!}$$

and we can add the fractions using the common denominator of $k!(m-k+1)!$. So

$$\begin{aligned}
\binom{n}{k} &= \frac{k}{k} \frac{m!}{(k-1)!((m+1)-k)!} + \frac{m-k+1}{m-k+1} \frac{m!}{k!(m-k)!} \\
&= \frac{k(m!) + (m+1)(m!) - k(m!)}{k!((m+1)-k)!} \\
&= \frac{(m+1)!}{k!((m+1)-k)!} = \frac{n!}{k!(n-k)!}
\end{aligned}$$

and so $P(n,k)$ is true.

So by induction we see that $P(n,k)$ is true for all natural numbers n, k with $0 \leq k \leq n$. □

4.3.3 Binomial Theorem

Related Content Standards

- (HSA.APR.4) Know and apply the Binomial Theorem for the expansion of $(x+y)^n$ in powers of x and y for a positive integer n, where x and y are any numbers, with coefficients determined for example by Pascal's Triangle.

Theorem 4.8 (Binomial Theorem)**.** *Let a, b be in a number system with the distribution property and $ab = ba$, and let $n \in \mathbb{N}$, then*

$$(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \cdots + \binom{n}{n-1}a^1 b^{n-1} + \binom{n}{n}a^0 b^n.$$

We need commutativity of a and b so that $b(a^k b^l) = a^k b^{l+1}$ in the proof below.

Proof. We will prove this using induction on the exponent. So for a, b and for $n \in \mathbb{N}$ we let $P(n)$ be the statement that

$$(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \cdots + \binom{n}{n-1}a^1 b^{n-1} + \binom{n}{n}a^0 b^n.$$

So $P(0)$ is the statement that $(a+b)^0 = 1$, which is defined to be true.

If we assume that for some integer m that $P(m)$ is true, we see that

$$\begin{aligned}
(a+b)^{m+1} =& (a+b)\cdot(a+b)^m \\
=& (a+b)\left(\binom{m}{0}a^m b^0 + \binom{m}{1}a^{m-1}b^1 + \cdots + \binom{m}{m-1}a^1 b^{m-1} + \binom{m}{m}a^0 b^m\right) \\
=& \left(\binom{m}{0}a^{m+1}b^0 + \binom{m}{1}a^m b^1 + \cdots + \binom{m}{m-1}a^2 b^{m-1} + \binom{m}{m}a^1 b^m\right) \\
& + \left(\binom{m}{0}a^m b^1 + \binom{m}{1}a^{m-1}b^2 + \cdots + \binom{m}{m-1}a^1 b^m + \binom{m}{m}a^0 b^{m+1}\right) \\
=& \binom{m}{0}a^{m+1}b^0 + \left(\sum_{k=1}^{m}\left(\binom{m}{k} + \binom{m}{k-1}\right)a^{(m+1)-k}b^k\right) + \binom{m}{m}a^0 b^{m+1} \\
=& \binom{m}{0}a^{m+1}b^0 + \left(\sum_{k=1}^{m}\binom{m+1}{k}a^{(m+1)-k}b^k\right) + \binom{m}{m}a^0 b^{m+1} \\
=& \binom{m+1}{0}a^{m+1}b^0 + \binom{m+1}{1}a^m b^1 + \cdots + \binom{m+1}{m}a^1 b^m + \binom{m+1}{m+1}a^0 b^{m+1}
\end{aligned}$$

since $\binom{m}{0} = \binom{m+1}{0} = 1$ and $\binom{m}{m} = \binom{m+1}{m+1} = 1$. So $P(m+1)$ is true.

Therefore, by induction, $P(n)$ is true for all $n \in \mathbb{N}$. ∎

For the first few values of n we see that we have the following:

$$\begin{array}{lccccccccccc}
(a+b)^0 = & & & & & & 1 & & & & & \\
(a+b)^1 = & & & & & 1a & + & 1b & & & & \\
(a+b)^2 = & & & & 1a^2 & + & 2ab & + & 1b^2 & & & \\
(a+b)^3 = & & & 1a^3 & + & 3a^2b & + & 3ab^2 & + & 1b^3 & & \\
(a+b)^4 = & & 1a^4 & + & 4a^3b & + & 6a^2b^2 & + & 4ab^3 & + & 1b^4 & \\
(a+b)^5 = & 1a^5 & + & 5a^4b & + & 10a^3b^2 & + & 10a^2b^3 & + & 5ab^4 & + & 1b^5
\end{array}$$

4.3.4 Exercises

1. Create divisibility tests in Base 12 that correspond to those given in this section for Base 10.

4.4 Rational Numbers

Since the only integers that have a multiplicative inverse are 1 and -1, it would be beneficial to expand our number system to include symbols that can operate as multiplicative inverses. In this process we allow the ordered pairs in $\mathbb{Z} \times (\mathbb{Z} - \{0\})$ to be written as a fraction,

$$\frac{a}{b} \cong (a, b).$$

But we also want the set to be closed under addition and multiplication. This results in a set that contains all possible rational expressions of integers whose denominator is non-zero. Therefore we create the set

$$\widehat{\mathbb{Q}} = \left\{ \frac{m}{n} \middle| m, n \in \mathbb{Z}, n \neq 0 \right\} \cong \mathbb{Z} \times (\mathbb{Z} - \{0\}).$$

Then, similar to our process of constructing the integers from the natural numbers, we define a relation of two rational expressions by

$$\frac{a}{b} \sim \frac{c}{d} \Leftrightarrow ad = bc.$$

We will now verify that this relation of rational expressions represents an equivalence relation.

Reflexive

One can see that this relation is reflexive, since if $\frac{a}{b}$, with $b \neq 0$, is a rational expression of integers that $ab = ab$ and so $\frac{a}{b}$ is equivalent to itself.

Symmetric

Since equality of integers is a symmetric relationship we have that for any integers a, b, c, d with $ad = bc$ that $bc = ad$ and so this equivalence of rational expressions is symmetric.

Transitive

The more challenging property of equivalence relations to prove is the transitive property. To prove transitivity, we assume that we have three rational expressions,

$$\frac{a}{b}, \frac{c}{d}, \text{ and } \frac{m}{n},$$

with b, d, and n all non-zero, with

$$\frac{a}{b} \sim \frac{c}{d} \quad \text{and} \quad \frac{c}{d} \sim \frac{m}{n}.$$

From the definition of equivalence of rational expressions we have that $ad = bc$ and $cn = md$. From the properties of multiplication of integers we have that

$$(ad) \cdot (cn) = (bc) \cdot (md),$$

or equivalently,

$$ancd = bmcd.$$

This can then be rewritten as

$$(an - bm) \cdot (cd) = 0.$$

Since the integers have the property that the product of two non-zero integers is also non-zero, we know that either $an - bm = 0$ or $cd = 0$. Since d is assumed to be non-zero, $cd = 0$ would imply that $c = 0$. This would then imply that both ad and md are zero, since $ad = bc$ and $cn = md$. Since $d \neq 0$ we would then have that $a = 0$ and $m = 0$, and consequently that $an = bm = 0$. Therefore, $an - bm = 0$ independently of $cd = 0$. Therefore, we have that $an = bm$ and that $\frac{a}{b} \sim \frac{m}{n}$, implying that equivalence of rational expressions is transitive.

Since this relation on rational expressions is an equivalence relation we should explore what the equivalence classes of these rational expressions look like.

We will first look at the rational expressions that are equivalent to $\frac{0}{1}$. We see that

$$\left[\frac{0}{1}\right] = \left\{ \frac{a}{b} \,\middle|\, b \neq 0,\ 0 \cdot b = 1 \cdot a \right\} = \left\{ \frac{0}{b} \,\middle|\, b \neq 0 \right\}$$

Similarly,

$$\left[\frac{1}{1}\right] = \left\{ \frac{a}{b} \,\middle|\, b \neq 0,\ 1 \cdot b = 1 \cdot a \right\} = \left\{ \frac{a}{a} \,\middle|\, a \neq 0 \right\}.$$

We can follow this process and see that we can embed the integers into this set of equivalence classes using the correspondence of

$$n \in \mathbb{Z} \quad \leftrightarrow \quad \left[\frac{n}{1}\right].$$

4.4.1 Operations

Now that we understand the set of equivalence classes of rational expressions, we need to define our operations of addition and multiplication on this set and verify that these definitions are well-defined.

Definition 4.3. Let $\left[\frac{a}{b}\right]$ and $\left[\frac{c}{d}\right]$ be two equivalence classes of rational expressions with $a, b, c, d \in \mathbb{Z}$ and $b, d \neq 0$. We define addition and multiplication as

$$\left[\frac{a}{b}\right] + \left[\frac{c}{d}\right] := \left[\frac{ad + bc}{bd}\right] \quad \text{and} \quad \left[\frac{a}{b}\right] \times \left[\frac{c}{d}\right] := \left[\frac{ac}{bd}\right].$$

Since b and d are non-zero integers, the no non-zero zero divisors property tells us that $bd \neq 0$ and so the operations of addition and multiplication are closed. We can make sure that they operate appropriately on the equivalence classes by making sure that different representations for the original equivalence classes of rational expression generate a sum and product that are elements of the same equivalence class. The process is similar to the same process that was done with the integers. Since we did the proof for addition being well-defined on the integers, we will do the proof that multiplication is well-defined for the rational numbers.

Let $\frac{a}{b} \sim \frac{a'}{b'}$ and $\frac{c}{d} \sim \frac{c'}{d'}$. Using the definition of the equivalence classes, we have that

$$a \cdot b' = b \cdot a' \quad \text{and} \quad c \cdot d' = d \cdot c'.$$

Therefore,

$$\begin{aligned}
(bd) \cdot (a'c') &= (ba') \cdot (dc') && \text{by the commutative property of multiplication} \\
&= (ab') \cdot (dc') && \text{since } ab' = ba' \\
&= (ab') \cdot (cd') && \text{since } cd' = dc' \\
&= (ac) \cdot (b'd') && \text{by the commutative property of multiplication}
\end{aligned}$$

and so we have that $\frac{ac}{bd} \sim \frac{a'c'}{b'd'}$.

So we have constructed a new number system that we call the rational numbers, $\mathbb{Q}$. In order to simplify our notation, we will no longer write each rational number as an equivalence class. We will instead just keep in mind the equivalence class structure and use any of the possible elements of the equivalence class to represent the rational number. We will also often write rational expressions of the form $\frac{n}{1}$ as the corresponding integer n. We will also write

$$\frac{-a}{b}, \quad \frac{a}{-b}, \text{ and } \quad -\frac{a}{b}$$

as equivalent expressions.

Related Content Standards

- (7.NS.2) Apply and extend previous understandings of multiplication and division and of fractions to multiply and divide rational numbers.
  a. Understand that multiplication is extended from fractions to rational numbers by requiring that operations continue to satisfy the properties of operations, particularly the distributive property, leading to products such as $(-1)(-1) = 1$ and the rules for multiplying signed numbers. Interpret products of rational numbers by describing real-world contexts.
  b. Understand that integers can be divided, provided that the divisor is not zero, and every quotient of integers (with non-zero divisor) is a rational number. If p and q are integers, then $-(p/q) = (-p)/q = p/(-q)$. Interpret quotients of rational numbers by describing real world contexts.
  c. Apply properties of operations as strategies to multiply and divide rational numbers.

4.4.2 Order

We can define an order on $\mathbb{Q}$ by first defining the relationship between a rational expression and 0, built on the order on the integers. We say that for a rational expression $\frac{a}{b}$, with $b \neq 0$ that

$$\left(\frac{a}{b} < 0 \Leftrightarrow ab < 0\right), \quad \left(\frac{a}{b} = 0 \Leftrightarrow ab = 0\right), \text{ and } \quad \left(\frac{a}{b} > 0 \Leftrightarrow ab > 0\right).$$

From the no non-zero divisors property of the integers, we know that one, and only one, of the three statements can be true for a single rational expression. If $\frac{c}{d} \sim \frac{a}{b}$, we use that $ad = bc$ and

$$(ad)^2 = (ad)(ad) = (ad)(bc) = (ab)(cd)$$

to see that ab and cd have the same relationship to 0. So this ordering with respect to 0 is well-defined.

We can then extend the ordering to the entire set of rational numbers by defining

$$\left(\frac{a}{b} < \frac{c}{d} \Leftrightarrow \left(\frac{a}{b} + \frac{-c}{d}\right) < 0\right), \quad \left(\frac{a}{b} = \frac{c}{d} \Leftrightarrow \left(\frac{a}{b} + \frac{-c}{d}\right) = 0\right),$$
$$\text{and} \quad \left(\frac{a}{b} > \frac{c}{d} \Leftrightarrow \left(\frac{a}{b} + \frac{-c}{d}\right) > 0\right).$$

If we have rational numbers $\frac{a}{b}$, $\frac{m}{n}$, and $\frac{p}{q}$ with

$$\frac{a}{b} < \frac{m}{n}, \quad \text{and} \quad \frac{m}{n} < \frac{p}{q},$$

then

$$\frac{m}{n} + \frac{-a}{b} > 0, \quad \frac{p}{q} + \frac{-m}{n} > 0$$

and so

$$\frac{p}{q} + \frac{-a}{b} = \frac{p}{q} + \frac{-m}{n} + \frac{m}{n} + \frac{-a}{b} > 0$$

and this relation satisfies the properties of an order and we can extend the properties of inequalities from the integers.

Theorem 4.9. *Let $r, s, t \in \mathbb{Q}$. Then*

$$(1)\ (r = s) \Leftrightarrow (r + t = s + t), \quad (2)\ (r < s) \Leftrightarrow (r + t < s + t),$$

$$(3)\ (r = s) \Leftrightarrow (r \times t = s \times t),\ \textit{with } t \neq 0 \quad \textit{and} \quad (4)\ (r < s) \Leftrightarrow (r \times t < s \times t),\ \textit{with } t > 0.$$

Proof. In order to get the idea of the proof we will prove parts (1) and (4) and leave the proofs of the other two parts to the reader.

Let $r = \frac{a}{b}$, $s = \frac{c}{d}$, and $t = \frac{m}{n}$ be rational numbers with $a, b, c, d, m, n \in \mathbb{Z}$ and $b, d, n \neq 0$.

Part (1). Then

$$(r = s) \Leftrightarrow (ad = bc) \Leftrightarrow \left((ad)n^2 = (bc)n^2\right) \Leftrightarrow ((an + bm)(dn) = (cn + dm)(bn))$$

since d and n are non-zero. We then see that

$$((an + bm)(dn) = (cn + dm)(bn)) \Leftrightarrow \left(\frac{an + bm}{bn} = \frac{cn + dm}{dn}\right) \Leftrightarrow (r + t = s + t)$$

by the definition of equality of rational numbers and properties of the operations on the integers.

Part (4). We can assume without any loss of generality that $b, d, n > 0$ by changing the signs in the numerators, if needed. Then we know that

$$(r < s) \Leftrightarrow ((s - r) > 0) \Leftrightarrow \left(\frac{cb - ad}{bd} > 0\right) \Leftrightarrow ((cb - ad)(bd) > 0) \Leftrightarrow ((cb - ad) > 0)$$

since $bd > 0$.

Similarly,

$$(r \times t < s \times t) \Leftrightarrow \left(\frac{am}{bn} < \frac{cm}{dn}\right) \Leftrightarrow \left(\frac{bcmn - admn}{bdn^2} > 0\right).$$

Since $b, d, n > 0$, we know that this is equivalent to

$$((bcmn - admn) > 0) \Leftrightarrow ((cb - ad)(mn) > 0) \Leftrightarrow ((cb - ad) > 0)$$

since $t > 0$ and $n > 0$ implies that $m > 0$ and $mn > 0$.

Therefore we have that

$$(r < s) \Leftrightarrow (r \times t < s \times t).$$

□

Related Content Standards

- (7.NS.1) Apply and extend previous understandings of addition and subtraction to add and subtract rational numbers; represent addition and subtraction on a horizontal or vertical number line diagram.
  a. Describe situations in which opposite quantities combine to make 0. For example, a hydrogen atom has 0 charge because its two constituents are oppositely charged.
  b. Understand $p+q$ as the number located a distance $|q|$ from p, in the positive or negative direction depending on whether q is positive or negative. Show that a number and its opposite have a sum of 0 (are additive inverses). Interpret sums of rational numbers by describing real-world contexts.
  c. Understand subtraction of rational numbers as adding the additive inverse, $p - q = p + (-q)$. Show that the distance between two rational numbers on the number line is the absolute value of their difference, and apply this principle in real-world contexts.
  d. Apply properties of operations as strategies to add and subtract rational numbers.

4.4.3 Algebraic Properties

Since the rational numbers were constructed from the integers, they maintain many of the corresponding properties of the integers. So if $\frac{a}{b}$, $\frac{m}{n}$, and $\frac{p}{q}$ are elements of $\mathbb{Q}$.

Table 4.3: Properties of Rational Numbers

		Property
$\frac{a}{b} + \frac{m}{n} \in \mathbb{Q}$	$\frac{a}{b} \cdot \frac{m}{n} \in \mathbb{Q}$	Closure
$\frac{a}{b} + \left(\frac{m}{n} + \frac{p}{q}\right) = \left(\frac{a}{b} + \frac{m}{n}\right) + \frac{p}{q}$	$\frac{a}{b} \cdot \left(\frac{m}{n} \cdot \frac{p}{q}\right) = \left(\frac{a}{b} \cdot \frac{m}{n}\right) \cdot \frac{p}{q}$	Associative Property
$\frac{a}{b} + \frac{m}{n} = \frac{m}{n} + \frac{a}{b}$	$\frac{a}{b} \cdot \frac{m}{n} = \frac{m}{n} \cdot \frac{a}{b}$	Commutative Property
$\frac{a}{b} + 0 = \frac{a}{b}$	$\frac{a}{b} \cdot 1 = \frac{a}{b}$	Identities
$\frac{a}{b} \cdot \left(\frac{m}{n} + \frac{p}{q}\right) = \left(\frac{a}{b} \cdot \frac{m}{n}\right) + \left(\frac{a}{b} \cdot \frac{p}{q}\right)$	$\left(\frac{a}{b} + \frac{m}{n}\right) \cdot \frac{p}{q} = \left(\frac{a}{b} \cdot \frac{p}{q}\right) + \left(\frac{m}{n} \cdot \frac{p}{q}\right)$	Distributive Property
$\frac{a}{b} + \left(-\frac{a}{b}\right) = 0$	If $a, b \neq 0$, $\frac{a}{b} \cdot \left(\frac{b}{a}\right) = 1$	Inverses
If $\frac{a}{b}, \frac{m}{n} > 0$, then $\frac{a}{b} \cdot \frac{m}{n} > 0$.	If $\frac{a}{b}, \frac{m}{n} < 0$, then $\frac{a}{b} \cdot \frac{m}{n} > 0$.	
$\frac{a}{b} \cdot \frac{m}{n} = 0$ if, and only if, $\frac{a}{b} = 0$, $\frac{m}{n} = 0$, or both		No zero divisors

4.4.4 Properties of Exponents

Using our definition of exponents from Section 4.1, we will expand the number system for the base from the integers to the rational numbers.

If $\dfrac{p}{q}$ (with $q \neq 0$) is a rational number, for any natural number n we define

$$\left(\frac{p}{q}\right)^n := \frac{p^n}{q^n}.$$

Such a definition makes sense because we know from Section 4.1 that p^n and q^n are both well-defined integers, that $q^n \neq 0$, and the ratio of such integers in a rational number.

We can then extend the number system for the exponents from the natural numbers to the integers for $p, q \neq 0$ by defining

$$\left(\frac{p}{q}\right)^{-n} = \left(\frac{q}{p}\right)^n$$

for all $n \in \mathbb{N}$. In the case that $p = 0$, this definition would not be well-defined and so we define $0^n = 1$ for all $n \in \mathbb{Z}$.

Related Content Standards

- (8.EE.1) Know and apply the properties of integer exponents to generate equivalent numerical expressions.

With the extended definition of exponents we still have for all rational numbers, a, that $a^0 = 1$ and $a^1 = a$. We also have from our definition that for $a \neq 0$ and $n \in \mathbb{Z}$,

$$a^{-n} = \frac{1}{a^n} = \left(\frac{1}{a}\right)^n.$$

For $m, n \in \mathbb{N}$ and integers p and q, with $q \neq 0$,

$$\left(\frac{p}{q}\right)^m \cdot \left(\frac{p}{q}\right)^n = \left(\frac{p^m}{q^m}\right) \cdot \left(\frac{p^n}{q^n}\right) = \frac{p^m \cdot p^n}{q^m \cdot q^n} = \frac{p^{(m+n)}}{q^{(m+n)}} = \left(\frac{p}{q}\right)^{(m+n)}.$$

This can then be extended to $m, n \in \mathbb{Z}$ with

- [Case 1:] If $p = 0$, then $\frac{p}{q} = 0$ and
$$0^m \cdot 0^n = 0 \cdot 0 = 0 = 0^{(m+n)}.$$
- [Case 2:] If $p \neq 0$, we let $m, n \in \mathbb{Z}$. If m and n are both non-negative, then we have the property given above. If m and n are both negative, then
$$\left(\frac{p}{q}\right)^m \cdot \left(\frac{p}{q}\right)^n = \left(\frac{q}{p}\right)^{-m} \cdot \left(\frac{q}{p}\right)^{-n} = \left(\frac{q}{p}\right)^{(-m-n)} = \left(\frac{p}{q}\right)^{m+n}.$$
If one is negative, m, and the other is non-negative, n, then
$$\left(\frac{p}{q}\right)^m \cdot \left(\frac{p}{q}\right)^n = \frac{p^m}{q^m} \cdot \frac{q^{-n}}{p^{-n}} = \frac{p^{m+n}}{q^{m+n}} = \left(\frac{p}{q}\right)^{m+n}.$$

Therefore, we have that for all $a \in \mathbb{Q}$ and for all $m, n \in \mathbb{Z}$ that $a^m \cdot a^n = a^{m+n}$.

One can make similar arguments to extend the remainder of Theorem 4.4 so that $(ab)^n = a^n \cdot b^n$ for each $n \in \mathbb{Z}$ and $(a^m)^n = a^{mn}$ for each $m, n \in \mathbb{Z}$.

Since the rational numbers include values between integers, it makes sense to explore the relationships between exponents and the order of the rational numbers given by inequalities. If $a \in \mathbb{Q}$ with $a > 1$, then we can write a as $\frac{p}{q}$ for some integers p and q with $0 < q < p$, and from Theorem 4.4 we have $0 < q^n < p^n$ for any integer $n \geq 1$. So we can conclude that $a^n > 1$. If n is a negative integer, then $0 < q^{-n} < p^{-n}$ and so $a^{-n} > 1$, and hence $a^n < 1$.

If a and b are rational numbers with $0 < a < b$, then $\frac{a}{b} \in \mathbb{Q}$ with $0 < \frac{a}{b} < 1$ so $0 < \left(\frac{a}{b}\right)^n < 1$ and $0 < a^m < b^m$.

We combine all of the results into the following theorem.

Theorem 4.10. *Let a and b be rational numbers.*

1. $a^0 = 1$ *and* $a^1 = a$
2. *If* $a > 1$, *then* $a^n > 1$ *for all integers* $n > 0$, *and if* $a < 1$, *then* $0 < a^n < 1$ *for all integers* $n < 0$
3. $a^{-n} = \frac{1}{a^n}$, $a^m \cdot a^n = a^{m+n}$, $(ab)^n = a^n \cdot b^n$, *and* $(a^m)^n = a^{mn}$ *for each* $m, n \in \mathbb{Z}$
4. *If* $0 < a < b$ *and* $m \in \mathbb{N}$, *then* $0 < a^m < b^m$.
5. *If* $a > 1$ *and* $m < n$, *then* $a^m < a^n$.

4.4.5 Exercises

1. Prove that there exists a rational number between any two rational numbers. (If $a, b \in \mathbb{Q}$ with $a < b$, then there exists a $c \in \mathbb{Q}$ such that $a < c < b$.)
2. Let $x, y \in \mathbb{Q}$ such that $0 < x < y$. Prove that $0 < \frac{1}{y} < \frac{1}{x}$.
3. Look graphically at $\mathbb{Z} \times (\mathbb{Z} - \{0\})$, which is a different representation of $\widehat{\mathbb{Q}}$ defined at the beginning of this section.
   a. How would you graphically describe the equivalence classes of the relation
   $$(a, b) \sim (c, d) \Leftrightarrow ad = bc?$$
   b. How does this interact with the standards in Grades 7 and 8 regarding ratio and direct proportions?

4. Compare the definition of order of rational numbers in this section with other techniques of determining the order of rational numbers (i.e. cross-multiplying, finding lowest common denominator, 'butterfly method', etc.).

   a. What types of (mis)conceptions arise with each of these methods?
   b. Is there an order to presenting these various methods for determining the order of rational numbers that improves student understanding?

4.5 Representations of Rational Numbers

Each rational number can be represented in several different ways, with the most prominent being as a mixed number, a fraction, a decimal, or a percent. In this section we will discuss how to convert between these different representations and how to use other physical and graphical representations to understand and use rational numbers.

4.5.1 Decimal Representations

It is sometimes helpful to write rational numbers in a decimal representation, rather than fraction notation. This is particularly true in the comparison of two rational numbers. One process for finding the decimal representation from a fraction is long division, but it is often simpler to use technology to find the decimal representation. (It also greatly reduces the amount of error involved.) However, it is important to have a deeper understanding of the decimal representation of rational numbers to verify that the result obtained from a calculator or computer is correct. For instance, if one uses a calculator to find $\frac{8}{17}$, a calculator with an 8 digit display will give the answer 0.4705882, and a calculator with a 16 digit display will give the answer 0.470588235294118. There is no apparent repetition in this decimal, leading to the question of if it repeats or not. It is the repeated use of the division algorithm that allows us to determine why a decimal representation of a rational number always repeats.

Related Content Standards

- (7.NS.2) Apply and extend previous understandings of multiplication and division and of fractions to multiply and divide rational numbers.

  d. Convert a rational number to a decimal using long division; know that the decimal form of a rational number terminates in 0s or eventually repeats.

We will state the division algorithm for integers here, but the proof is in Section 7.1.

Theorem 4.11 (The Division Algorithm for Integers)**.** *If a and b are integers with $b > 0$, then there are unique integers q and r such that*

$$a = bq + r, \quad \textit{with } 0 \leq r < b.$$

We will explain the process of long division and its relationship to the division algorithm to determine the decimal representation of rational numbers using an example.

In order to find the decimal expansion of $\frac{7}{12}$, we will first rewrite the rational expression as $\frac{1}{10} \cdot \frac{70}{12}$. We then note that $70 = 12 \cdot 5 + 10$ in the division algorithm. Thus,

$$\frac{7}{12} = \frac{1}{10} \cdot \frac{70}{12} = \frac{1}{10} \cdot \left(5 + \frac{10}{12}\right) = \frac{5}{10} + \frac{1}{10} \cdot \frac{10}{12}.$$

We continue this process with $\frac{10}{12}$ so that we have

$$\frac{7}{12} = \frac{5}{10} + \frac{1}{10^2}\left(\frac{100}{12}\right) = \frac{5}{10} + \frac{1}{10^2}\left(8 + \frac{4}{12}\right) = \frac{5}{10} + \frac{8}{10^2} + \frac{1}{10^2} \cdot \frac{4}{12}.$$

We continue the process, so that

$$\frac{7}{12} = \frac{5}{10} + \frac{8}{10^2} + \frac{1}{10^3} \cdot \frac{40}{12} = \frac{5}{10} + \frac{8}{10^2} + \frac{3}{10^3} + \frac{1}{10^3} \cdot \frac{4}{12}.$$

We notice that this remainder of 4 again tells us that we will get the same quotient again and again. Therefore, we see that $\frac{7}{12} = 0.58\overline{3}$.

As we can see with this process, a decimal will start repeating when a remainder in the division algorithm is repeated. Since the only possible remainders of a quotient by 12 are in the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$, we see that any decimal representation of a fraction whose denominator is 12 will repeat after at most 11 places.

Similarly, the decimal representation for any rational number will eventually repeat. This repetition may be of the digit 0, in which case we will call it a **finite decimal expansion**, or it may be a different number of digits, in which we will call it a **repeating decimal expansion**. In the case of $\frac{7}{12}$, we see that the repeating happens after a certain number of digits. These types of decimal representations will be called **delayed repeating decimal expansions**. One can determine the type of decimal expansion based on the prime factorization of the denominator, once the fraction is written in its simplified form.

Related Content Standards

- (8.NS.1) Know that numbers that are not rational are called irrational. Understand informally that every number has a decimal expansion; for rational numbers show that the decimal expansion repeats eventually, and convert a decimal expansion which repeats eventually into a rational number.

Now that we know that all rational numbers can be represented by decimal expansions that are finite, repeating, or delayed repeating, we will study the process of starting with a decimal representation and determining a fraction representation. We already know that if the decimal representation does not repeat that it does not correspond to a rational number.

If the number has a finite decimal expansion, then it can be converted to a fraction using the base ten representation. For instance the number 3627.854 can be rewritten as $\frac{3627854}{1000}$.

Let's look at a decimal expansion that is repeating like $0.\overline{35682}$. This could be written in the form

$$0.\overline{35682} = \frac{10^5}{10^5} 0.\overline{35682} = \frac{1}{10^5} \cdot 35682.\overline{35682} = \frac{1}{10^5}\left(35682 + 0.\overline{35682}\right).$$

We can rearrange this equation so that

$$10^5 \cdot 0.\overline{35682} - 0.\overline{35682} = 35682$$

or equivalently that

$$0.\overline{35682} = \frac{35682}{10^5 - 1}.$$

This same process can be used to find the fractional representation of any rational number with a repeating decimal expansion.

If a rational number has a delayed repeating decimal expansion, then one can rewrite the number in a way to generate a repeating decimal. For instance,

$$324.51\overline{89} = \frac{10^2}{10^2} \cdot 324.51\overline{89} = \frac{1}{10^2}\left(32451.\overline{89}\right) = \frac{1}{10^2}\left(32451 + \frac{89}{10^2 - 1}\right) = \frac{(32451 \cdot 99) + 89}{9900}.$$

4.5.2 Physical and Graphical Representations

Related Content Standards

- (6.NS.1) Interpret and compute quotients of fractions, and solve word problems involving division of fractions by fractions, e.g., by using visual fraction models and equations to represent the problem. For example, create a story context for $\frac{2}{3} \div \frac{3}{4}$ and use a visual fraction model to show the quotient; use the relationship between multiplication and division to explain that $\frac{2}{3} \div \frac{3}{4} = \frac{8}{9}$ because $\frac{3}{4}$ of $\frac{8}{9}$ is $\frac{2}{3}$. (In general, $\frac{a}{b} \div \frac{c}{d} = \frac{ad}{bc}$.) How much chocolate will each person get if 3 people share $\frac{1}{2}$ lb of chocolate equally? How many $\frac{3}{4}$-cup servings are in $\frac{2}{3}$ of a cup of yogurt? How wide is a rectangular strip of land with length $\frac{3}{4}$ mi and area $\frac{1}{2}$ square mi?

Consider the following problem from the Singapore Mathematics Curriculum:

> Jenny, Bob, and Paul shared a sum of money. Jenny received $\frac{2}{5}$ of the money, Bob received $\frac{1}{4}$ of the money and Paul received the rest of the money. If Paul received \$1.50 more than Bob, how much more money did Jenny receive than Paul.

While this is not a modeling problem, in that this type of situation would never be described in this way, it does help generate fluency in working with different representations of rational numbers. We will walk through one possible solution that does not use variables, but does build the conceptual understanding necessary for successful work in algebra.

The first thing that we notice is that we need a way to represent $\frac{2}{5}$ and $\frac{1}{4}$ of the same whole. As such, we need our whole item to be broken into 20 equal sized pieces.

We then note that Jenny received $\frac{2}{5}$ of the money, which is equivalent to $\frac{8}{20}$ of the money. So we will mark 8 of the pieces with a J. Similarly, Bob received $\frac{1}{4}$ of the money and so we will mark 5 pieces with a B.

J	J	J	J	J
J	J	J		
B	B	B	B	B

We then see that Paul received the equivalent of 7 blocks, while Bob received 5 blocks. Since Paul received \$1.50 more than Bob we know that 2 blocks is equivalent to \$1.50, making each block represent \$0.75. Since Jenny received the equivalent of 1 block more than Paul we see that she received \$0.75 more than Paul.

There are many other ways to solve this problem that also develop numerical fluency with rational numbers, and the more ways that students can represent these numbers and move between the representations, the more fluent they will be with the rational number system.

4.5.2.1 Fraction Strips

One common tool used to help students determine equivalent fraction representations is through the use of fraction strips. These are pieces of paper (all the same size), but marked off in different numbers of parts.

$\frac{1}{1}$											
$\frac{1}{2}$						$\frac{1}{2}$					
$\frac{1}{3}$				$\frac{1}{3}$				$\frac{1}{3}$			
$\frac{1}{4}$			$\frac{1}{4}$			$\frac{1}{4}$			$\frac{1}{4}$		

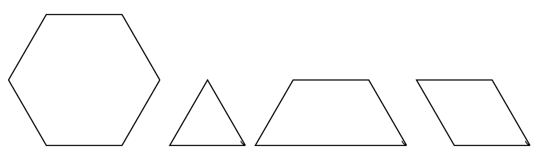
Then by setting the pieces of paper next to each other, students are able to notice equivalency of fractions. For instance, in the fraction strips below one can see that $\frac{1}{2}$ is equivalent to $\frac{2}{4}$.

One can also use fraction strips to represent the operation of division of fractions. For instance if we have the two fraction strips below we can use them to determine that $\frac{1}{2} \div \frac{1}{6} = 3$ because 3 of the $\frac{1}{6}$ regions can fit inside of the $\frac{1}{2}$ region.

$\frac{1}{1}$											
$\frac{1}{2}$						$\frac{1}{2}$					
$\frac{1}{3}$				$\frac{1}{3}$				$\frac{1}{3}$			
$\frac{1}{4}$			$\frac{1}{4}$			$\frac{1}{4}$			$\frac{1}{4}$		

4.5.2.2 Pattern Blocks

Another tool to help students work with fractions are pattern blocks. The pattern blocks most useful for many of the fractions that students use are the hexagon, triangle, trapezoid, and rhombus below.

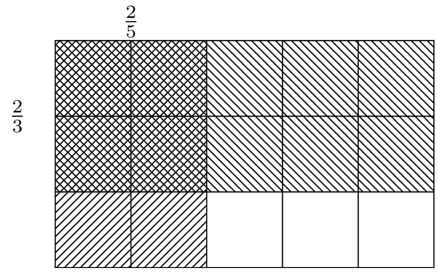


Students can see that 6 triangles, 2 trapezoids, or 3 rhombi fit inside of the hexagon. So we can represent $\frac{2}{3}$ as two rhombi, with one hexagon representing the whole. One can then see that fractions whose denominator is 2, 3, or 6 are easily represented using one hexagon as the whole. However, if we allow two hexagons to represent the whole, we can represent fractions with denominators of 2, 3, 4, 6, or 12. One can also think about $\frac{1}{2} \div \frac{1}{3}$ as how many rhombi fit inside of one trapezoid.

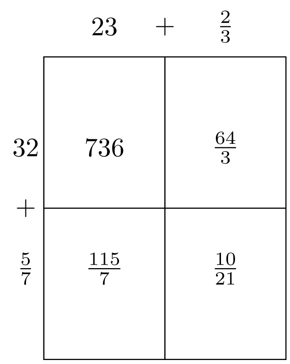
4.5.2.3 Area Model

The area model described in Section 4.2 can help students to visualize multiplication of fractions. For instance, one can find the product of $\frac{2}{5}$ and $\frac{2}{3}$ using the following diagram. One can then see that $\frac{4}{15}$ of the squares are shaded. So $\frac{2}{5} \times \frac{2}{3} = \frac{4}{15}$.

We are also able to build upon the area model to visualize the distribution property involved in the multiplication of mixed numbers. If we want to find $23\frac{2}{3} \times 32\frac{5}{7}$, then we have to remember the convention that $23\frac{2}{3} = 23 + \frac{2}{3}$ and $32\frac{5}{7} = 32 + \frac{5}{7}$. (Note how confusing this convention can be for students, particularly since in algebra two numbers next to each other represent multiplication instead of addition.) We can then



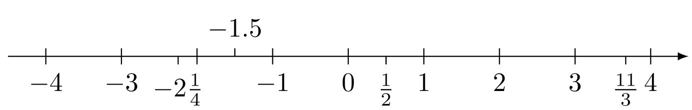
allow for an abstract area representation (the sizes are not proportional) to more easily use the distributive property.



There are many other tools that can help students understand rational numbers and their operations, but we will leave those for more pedagogically oriented sources.

4.5.2.4 Number Line and Plane

It is also important for students to use rational numbers on a number line (and the associated plane). In particular, the number line can help students understand inequalities, relationships between different numbers, and the absolute value of a number. Some key factors to keep in mind are to not only have fractions between 0 and 1, or only mixed numbers outside of that interval. Students need to use a variety of different numbers and representations together to gain fluency.



Related Content Standards

- (6.NS.7) Understand ordering and absolute value of rational numbers.
  a. Interpret statements of inequality as statements about the relative position of two numbers on a number line diagram.
  b. Write, interpret, and explain statements of order for rational numbers in real-world contexts.
  c. Understand the absolute value of a rational number as its distance from 0 on the number line; interpret absolute value as magnitude for a positive or negative quantity in a real-world situation.
  d. Distinguish comparisons of absolute value from statements about order.

4.5.3 Exercises

1. Explain, in a way that a middle school student can understand, how to determine the type of decimal expansion for any given fraction representation of a rational number and why it is that way. (i.e. when is the decimal expansion finite? when does it repeat? when is it delayed periodic? what do each of these mean?)

2. For each of the following, give an example of a rational number in the form of $\frac{p}{q}$ with the appropriate decimal representation and explain why it works.

   a. has a terminating decimal with 4 decimal places
   b. has a non-zero integer part and a delay of 3 decimal places before a periodic part with period of 3 digits.
   c. has a periodic decimal representation with 5 digits.

3. For each of the following, give the representation of the rational number in the form of $\frac{p}{q}$, with p and q relatively prime (no common factors):

   a. 0.156
   b. $0.23\overline{4}$
   c. $23.\overline{15}$
   d. 3.14159
   e. $2.5\overline{9}$
   f. 2016

4. Write an algorithm in the programming language of your choice to input a decimal representation of a rational number and output a fraction representation.

5. Classify each of the following real numbers as

   - a terminating decimal,
   - simple periodic decimal (note its period),
   - delayed periodic decimal (note its delay and period), or
   - non-periodic decimal

   a. $\frac{1}{6}$
   b. $\frac{1}{2}$
   c. $\frac{2}{14}$
   d. $\frac{3}{20}$
   e. $\frac{4}{63}$
   f. π
   g. $\sqrt{2}$
   h. $\sqrt{5}$
   i. $\frac{18}{40}$
   j. 0.75
   k. $0.83\overline{45}$

4.6 Real Numbers

One of the properties of the integers that does not extend to the rational numbers is the least-upper-bound property (3.4). In order to prove this, we first need to prove a series of small lemmas (a smaller version of a theorem that is used to prove a larger theorem).

Lemma 4.7. *If n is an integer, then n can be written as $2k$ for some integer k or $2j+1$ for some integer j, but cannot be written both ways. (In other words, an integer is even or odd, but not both.)*

Proof. We will use the technique of a proof by contradiction and assume that there is an integer n such that $n = 2k$ and $n = 2j + 1$ for some integers j and k. Then $2j = 2k + 1$ and so the equivalent equation $2(j - k) = 1$ is true. However, since 2 does not have a multiplicative inverse in the integers, $j - k$ cannot be an integer. This means that our assumption is false and our lemma is proven. $\square$

Lemma 4.8. *If n is an integer such that n^2 is even, then n is even.*

Proof. We will prove the contrapositive of the statement, which is logically equivalent to the original statement. Let n be an odd integer, i.e. $n = 2j + 1$ for some integer j. Then $n^2 = (2j + 1)^2 = 4j^2 + 4j + 1 = 2(2j^2 + 2j) + 1$ and so n^2 is odd. $\square$

Lemma 4.9. *There does not exist a rational number whose square is* 2.

Proof. We will again use the technique of a proof by contradiction. Assume that there is a rational number whose square is 2. We will denote the rational number by $\frac{a}{b}$ where a and b are not both even. (This can be accomplished by finding an equivalent representation without 2 as a common factor.) Since its square is 2, we have that $\left(\frac{a}{b}\right)^2 = 2$ and equivalently that $a^2 = 2b^2$. Since a^2 is even, the previous lemma tells us that a is even. So there exists an integer k such that $a = 2k$ and so $a^2 = 4k^2$. We then have that $4k^2 = 2b^2$ and so $b^2 = 2k^2$. This means that b^2 is even and consequently that b is even. This contradicts our original assumption and so no such rational number exits. $\square$

We will now look at the set

$$S = \left\{ \frac{a}{b} \in \mathbb{Q} \middle| \frac{a}{b} > 0 \text{ and } \left(\frac{a}{b}\right)^2 \geq 2 \right\}$$

which is bounded below in its definition and does not have a least element. Therefore, the rational numbers do not satisfy the least-upper-bound property.

Since the rational numbers do not satisfy the least-upper-bound property, the rational numbers have "holes" that need to be filled, and so are not complete. Building upon the work of Bertrand [1840] and in a parallel vein to Heine [1872], Dedekind [1872] developed a construction of the real numbers from the rational numbers using what are now referred to as Dedekind cuts using equivalence classes of such cuts. Another method of construction of the real numbers from the rational numbers follows the path of Cauchy [1821] and defines the real numbers as equivalence classes of Cauchy sequences of rational numbers. As each of these methods requires the development of a great deal of mathematical machinery, we will refer to other textbooks in analysis for the details. With Rudin [1976] providing an excellent treatment of Dedekind cuts in the Appendix of Chapter 1. Chapter I of Thurston [1956] provides an detailed treatment of the construction of the Cauchy numbers to generate the real numbers through equivalence classes.

Because the real numbers form a complete number system in that it satisfies the least-upper-bound property, for any bounded set of real numbers, S, we can define $\sup S$ to the be the least element of the set $\{x \in \mathbb{R} | x \geq a, \forall a \in S\}$ and $\inf S$ is the greatest element of the set $\{x \in \mathbb{R} | x \leq a, \forall a \in S\}$. These two attributes of a set play an important role in the development of calculus.

Since $\sqrt{2}$ is not rational, but we know it is a real number because of the completeness of the reals, and so there are real numbers that are not rational. We define these numbers to be **irrational**.

Definition 4.4. A real number, a, is called **rational** if there exist integers p and q such that $a = \frac{p}{q}$. A real number that is not rational is called **irrational**.

Related Content Standards

- (HSN.RN.3) Explain why the sum or product of two rational numbers is rational; that the sum of a rational number and an irrational number is irrational; and that the product of a non-zero rational number and an irrational number is irrational.

We have already shown that the sum and product of two rational numbers are rational, but we would like to explore how rational and irrational numbers interact with each other.

Theorem 4.12. *Let $u \in \mathbb{Q}$ and $v \in (\mathbb{R} - \mathbb{Q})$. Then $u + v \notin \mathbb{Q}$ and if $u \neq 0$, $uv \notin \mathbb{Q}$.*

Proof. When trying to prove that something does not satisfy a certain property, then it is best to use a proof by contradiction or to prove the contrapositive.

Let $u \in \mathbb{Q}$ and so there exist integers p and q such that $u = \frac{p}{q}$. Let $v \in (\mathbb{R} - \mathbb{Q})$. We will now prove that $u + v$ is not rational using a proof by contradiction. If $u + v$ is rational, then there exist integers m and n such that $u + v = \frac{m}{n}$. This means that

$$v = \frac{m}{n} - u = \frac{m}{n} - \frac{p}{q} = \frac{mq - np}{nq}$$

making it a rational number, contradicting the assumption that it is irrational.

The proof that uv is irrational is very similar. $\square$

One of the important properties to notice about the rational and irrational numbers is that they are dense in the real numbers in that between any two real numbers one can find both a rational and an irrational number.

Theorem 4.13 (Density of the Rational Numbers)**.** *Let $x, y \in \mathbb{R}$ be any two real numbers where $x < y$. Then there exists a rational number $q \in \mathbb{Q}$ such that $x < q < y$.*

Proof. Since $x < y$, we know that $y - x > 0$. So there exists a positive natural number $n \in \mathbb{N}$ such that $0 < \frac{1}{n} < y - x$. Equivalently, $0 < 1 < ny - nx$ and $nx + 1 < ny$. From properties of the integers we know that there exists an integer $N \in \mathbb{Z}$ such that $N - 1 \leq nx < N$, or equivalently $N \leq nx + 1 < N + 1$. Therefore,

$$nx < N \leq nx + 1 < ny$$

and so $x < \frac{N}{n} < y$. So $q = \frac{N}{n}$ is a rational number such that $x < q < y$. $\square$

Theorem 4.14 (Density of the Irrational Numbers)**.** *Let $x, y \in \mathbb{R}$ be any two real numbers where $x < y$. Then there exists an irrational number $v \in \mathbb{R} \setminus \mathbb{Q}$ such that $x < v < y$.*

Proof. Since x and y are real numbers, $\frac{x}{\sqrt{2}}$ and $\frac{y}{\sqrt{2}}$ are also real numbers such that $\frac{x}{\sqrt{2}} < \frac{y}{\sqrt{2}}$. By the density of the rational number, there exists a $q \in \mathbb{Q}$ such that

$$\frac{x}{\sqrt{2}} < q < \frac{y}{\sqrt{2}},$$

or equivalently $x < q\sqrt{2} < y$. Since $q \in \mathbb{Q}$ and $\sqrt{2}$ is irrational, $q\sqrt{2}$ is irrational and so if $v = q\sqrt{2}$, $v \in \mathbb{R} \setminus \mathbb{Q}$ such that $x < v < y$. $\square$

Since the operations of addition and multiplication on the real numbers are associative and commutative, for each natural number n we can define a^n recursively and we can prove the following properties using the same techniques as in Section 4.1, and Section 4.4 using limit ideas from the definition of the real numbers.

Theorem 4.15. *Let $a, b \in \mathbb{R}$ and let $m, n \in \mathbb{Z}$.*

- *$a^0 = 1$ and $a^1 = a$*
- *If $a > 1$ and $n > 0$, then $a^n > 1$. If $a > 1$ and $n < 0$, then $0 < a^n < 1$.*
- *$a^{-n} = \frac{1}{a^n}$*

- $a^m \cdot a^n = a^{m+n}$, $(ab)^n = a^n \cdot b^n$, *and* $\$\ (a^{m)}n = a\hat{\ }\{mn\}\$$
- *If* $0 < a < b$ *and* $m \in \mathbb{N}$, *then* $0 < a^m < b^m$.
- *If* $a > 1$ *and* $m < n$, *then* $a^m < a^n$.

With these properties, we are able to define define algebraic numbers, which is a number system between the rational numbers and the real numbers.

Definition 4.5. A real number, a, is called **algebraic** if

$$n_0 + n_1 a + n_2 a^2 + \cdots + n_m a^m = 0$$

for some integers $n_0, n_1, n_2, \ldots, n_m$.

For all rational numbers in the form $\frac{p}{q}$, we know that

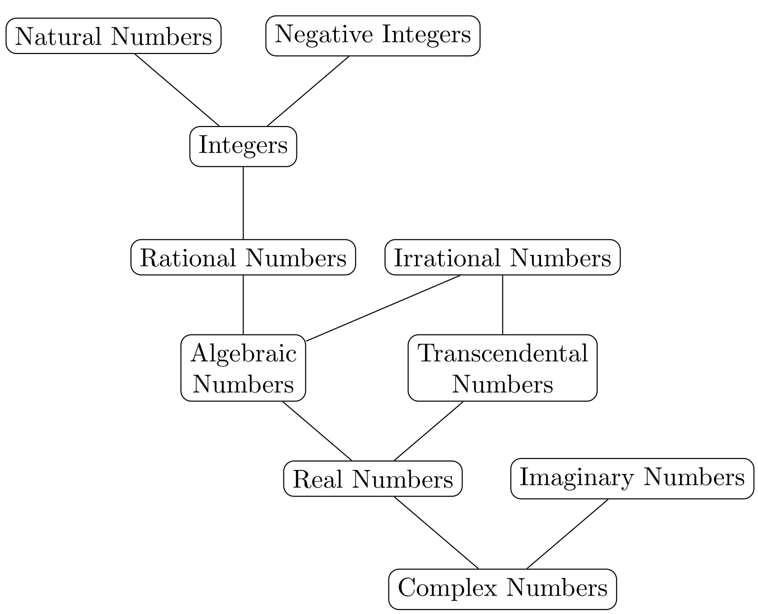
$$q \cdot \left(\frac{p}{q}\right)^1 + (-p) = 0$$

and so all rational numbers are algebraic numbers. We can also see that $\sqrt{2}$ is an algebraic number since $(\sqrt{2})^2 - 2 = 0$. We say that the algebraic numbers are a number system because they include the additive identity, 0; the multiplicative identity, 1; and are closed under addition and multiplication.

We will prove in Section 5.4 that there are an infinite number of real numbers that are not algebraic. We will call any real number that is not algebraic a **transcendental** number. Two such examples of transcendental numbers are e and π.

In the early 1930's, Gel'fond [1934] and Schneider [1935] independently solved Hilbert's seventh problem from his 1900 address to the International Congress of Mathematics showing that if a and b are algebraic numbers with $a \neq 0, 1$ and b irrational that a^b is a transcendental number. Since we have such a limited knowledge of these transcendental numbers, it is evident that the continuum of the real numbers is a very deep and complicated concept that we are only beginning to understand.

We include a graphical representation of how these different sets of numbers fit together. All of these sets of numbers are subsets of the complex numbers, and a set above another set represents an inclusion relationship.



4.6.1 Properties of Exponents

Let $a > 0$ be a real number. For each integer $n > 0$, we define

$$S_n = \{x \in \mathbb{Q} | x > 0 \text{ and } x^n > a\}$$

and so we can define

$$\sqrt[n]{a} = a^{\frac{1}{n}} := \inf S_n.$$

Related Content Standards

- (8.EE.2) Use square root and cube root symbols to represent solutions to equations of the form $x^2 = p$ and $x^3 = p$, where p is a positive rational number. Evaluate square roots of small perfect squares and cube roots of small perfect cubes. Know that $\sqrt{2}$ is irrational.

If $a > 0$ is a real number, m, n, p, q are integers, $n > 0$, and $q > 0$ such that $\frac{m}{n} = \frac{p}{q} = r$, we have that $mq = pn$. So

$$\left(\left(a^{\frac{1}{n}}\right)^m\right)^{nq} = a^{mq} = a^{pn} = \left(\left(a^{\frac{1}{q}}\right)^p\right)^{nq}.$$

This means that we can define $a^r := (a^m)^{\frac{1}{n}}$.

Related Content Standards

- (HSN.RN.1) Explain how the definition of the meaning of rational exponents follows from extending the properties of integer exponents to those values, allowing for a notation for radicals in terms of rational exponents.
- (HSN.RN.2) Rewrite expressions involving radicals and rational exponents using the properties of exponents.

We can now prove various exponential properties of real numbers with rational exponents. In order to prove a generalization of Theorem 4.15 to rational exponents we will only need to prove the corresponding properties for rational numbers of the form $\frac{1}{n}$ and it directly follows from $(a^n = b^n) \Leftrightarrow a = b$ for real numbers a and b and positive integers n. As such, we will omit the majority of these proofs from this text. However, will include one lemma and proof in order to understand the flavor of the proofs.

Lemma 4.10. *Let $a \in \mathbb{R}^+$. For all $r, s \in \mathbb{Q}$, $a^{r+s} = a^r \cdot a^s$.*

Proof. Let $a \in \mathbb{R}^+$ and let $r, s \in \mathbb{Q}$ such that $r = \frac{m}{n}$ and $s = \frac{p}{q}$ for some integers m, n, p, and q.

Since $q \neq 0$ and $n \neq 0$,

$$\left(a^{r+s}\right)^{qn} = \left(a^{\frac{mq+pn}{qn}}\right)^{qn} = a^{(mq+pn)}$$

and

$$(a^r \cdot a^s)^{qn} = \left(a^{\frac{mqn}{n}}\right) \cdot \left(a^{\frac{pqn}{q}}\right) = a^{(mq+pn)}.$$

Therefore, $a^{r+s} = a^r \cdot a^s$. $\square$

For a real number y, define

$$S_y = \{a^r | r \leq y\}.$$

Furthermore, define

$$a^y = \sup S_y.$$

Using arguments from analysis and the properties of limits, one can extend the previous properties of exponents to real valued exponents as given in Theorem 4.16. Due to the mathematical machinery necessary for these proofs, we will refer the reader to an analysis textbook.

Theorem 4.16. *Let a and b be positive real numbers and let $x, y \in \mathbb{R}$.*

- $a^0 = 1$ *and* $a^1 = a$
- $a^x > 1$ *for all* $x > 0$, *and* $0 < a^x < 1$ *for all* $x < 0$
- $a^{-x} = \frac{1}{a^x}$
- $a^x \cdot a^y = a^{x+y}$, $(ab)^x = a^x \cdot b^x$, *and* $(a^x)^y = a^{xy}$
- *If* $0 < a < b$ *and* $x > 0$, *then* $a^x < b^x$.
- *If* $a > 1$ *and* $x < y$, *then* $a^x < a^y$.

4.6.2 Exercises

1. If $n = 2$, $a^{\frac{1}{2}}$ is often written as $\sqrt{2}$, but for $n = \{3, 4, 5, \ldots\}$, $a^{\frac{1}{n}}$ is written as $\sqrt[n]{a}$. This difference in notation can cause many problems for students. What are some other notation differences in the secondary curriculum that have similar situations?
2. What can be said about the sum and product of two irrational numbers?
3. Suppose that $a \in \mathbb{R}$ with $a > 0$. If n is an even positive integer, show that there are exactly two real numbers that are solutions to the equation $x^n = a$.
4. Show that $\sqrt{2} + \sqrt{3}$ is an algebraic number by finding a polynomial for which it is a zero.
5. Write out a detailed proof that $\sqrt{3}$ is irrational using the techniques of the proof that $\sqrt{2}$ is irrational. For which numbers can this process be generalized?

4.7 Complex Numbers

Unlike all of the previous number system constructions, in order to construct the complex numbers, $\mathbb{C}$, we do not have to work through equivalence classes. Instead, we look at the set

$$\mathbb{R} \times \mathbb{R} = \{(x, y) | x, y \in \mathbb{R}\}$$

and define the addition and multiplication operations as

$$(a, b) + (c, d) = (a + c, b + d) \quad \text{and} \quad (a, b) \cdot (c, d) = (ac - bd, ad + bc).$$

We can see that elements of the form $(a, 0)$, $\mathbb{R} \times \{0\}$, act just like the real numbers in that

$$(a, 0) + (b, 0) = (a + b, 0) \quad \text{and} \quad (a, 0) \cdot (b, 0) = (ab, 0)$$

and so we can think of these as being the real numbers inside of the complex numbers.

Elements of the form $(0, a)$, $\{0\} \times \mathbb{R}$, operate very differently, particularly with regard to multiplication, since

$$(0, a) + (0, b) = (0, a + b) \quad \text{and } (0, a) \cdot (0, b) = (-ab, 0).$$

4.7.1 Rectangular Representation

If we define scalar multiplication on this set by $k \cdot (a, b) = (ka, kb)$, we see that the elements of the set $\mathbb{R} \times \mathbb{R}$ operate as a vector space in that for all $(a, b) \in \mathbb{R} \times \mathbb{R}$, $(a, b) = a \cdot (1, 0) + b \cdot (0, 1)$. If we then replace $(1, 0)$ with its corresponding element of 1 in $\mathbb{R}$ and label $0, 1$ with a new symbol that we call i, we have the complex numbers,

$$\mathbb{C} = \{a + bi | a, b \in \mathbb{R}\}$$

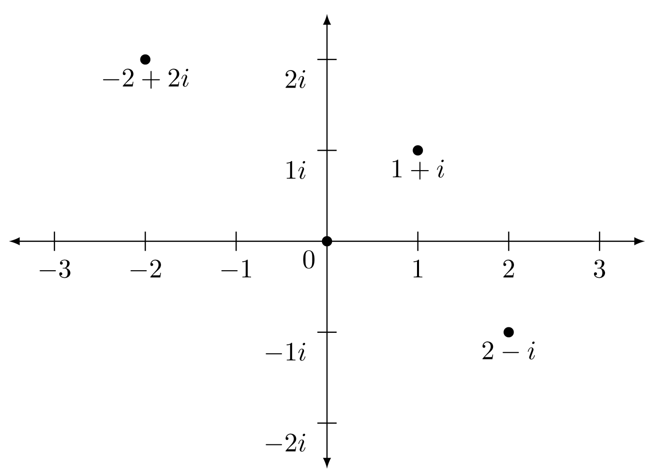
with addition and multiplication defined by

$$(a + bi) + (c + di) = (a + c) + (b + d)i \quad \text{and} \quad (a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i.$$

Related Content Standards

- (HSN.CN.1) Know there is a complex number i such that $i^2 = -1$, and every complex number has the form $a + bi$ with a and b real.
- (HSN.CN.2) Use the relation $i^2 = -1$ and the commutative, associative, and distributive properties to add, subtract, and multiply complex numbers.

In the eighth chapter of the greatest work of mathematics in the 18^{th} century, possibly even in the modern era, Leonard Euler [1748] described a geometric, algebraic, and analytic connection between ways of representing complex numbers that enabled many of the mathematical and scientific discoveries of the past 350 years. What Euler recognized is that in addition to representing complex numbers as $a + bi$, one can describe the complex numbers geometrically as points on the plane where the real part of the complex number provides the horizontal component and the imaginary part of the complex number provides the vertical component.



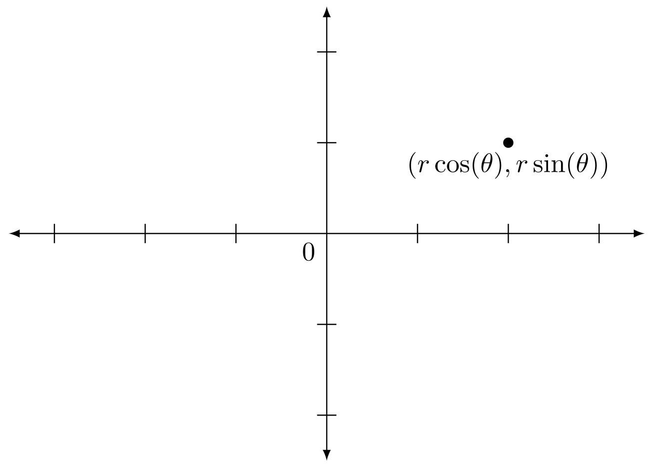
4.7.2 Complex Conjugation and Modulus

Related Content Standards

- (HSN.CN.4) Represent complex numbers on the complex plane in rectangular and polar form (including real and imaginary numbers), and explain why the rectangular and polar forms of a given complex number represent the same number.

Since points on the plane can also be described by the distance from the origin and the angle made with the positive horizontal axis, complex numbers can also be described by a magnitude and an angle. When we combine that with knowledge of trigonometric functions we have that every complex number can be written in the form

$$a + bi = r(\cos(\theta) + i\sin(\theta))$$



If we look at the correspondence of $a = r\cos(\theta)$ and $b = r\sin(\theta)$, we can derive formulas to solve for r and θ in terms of a and b,

$$r = \sqrt{a^2 + b^2} \quad \text{and} \quad \theta = \begin{cases} \arctan\left(\frac{b}{a}\right) & \text{if } a \geq 0 \\ \arctan\left(\frac{b}{a}\right) + \pi & \text{if } a < 0 \end{cases}.$$

In addition to this polar representation of a complex number, Euler defined the **complex conjugate** of a number $z = a + bi$ to be $\overline{z} = a - bi$. Geometrically, the complex conjugate of a number is its reflection across the horizontal axis. We also see that if we multiply a complex number $z = a + bi$ by its complex conjugate we have $(z\overline{z}) = a^2 + b^2 = r^2$, which is the square of the **modulus**, $|z|^2$, or distance from 0.

Notice that the modulus of a complex number is a real number. This makes it easier to divide by complex numbers. Using the complex conjugate we can write

$$\frac{z}{w} = \frac{z}{w} \cdot \frac{\overline{w}}{\overline{w}} = \frac{z\overline{w}}{|w|^2} \quad \text{or} \quad \frac{a+bi}{c+di} = \frac{(a+bi)(c-di)}{c^2+d^2} = \left(\frac{ac+bd}{c^2+d^2}\right) + \left(\frac{bc-ad}{c^2+d^2}\right) i$$

for complex numbers $z = a + bi$ and $w = c + di$.

This modulus of a complex number generates a distance between two complex numbers, $|z_2 - z_1|$, in a similar manner to absolute value measuring the distance between two real numbers, $|x_2 - x_1|$.

Related Content Standards

- (HSN.CN.3) Find the conjugate of a complex number; use conjugates to find moduli and quotients of complex numbers.
- (HSN.CN.5) Represent addition, subtraction, multiplication, and conjugation of complex numbers geometrically on the complex plane; use properties of this representation for computation.
- (HSN.CN.6) Calculate the distance between numbers in the complex plane as the modulus of the difference, and the midpoint of a segment as the average of the numbers at its endpoints.

4.7.3 Euler's Equation

Euler's [1748] greatest discovery that one can use infinite series to find a strong connection between exponential functions and trigonometric functions and that

$$e^{x+yi} = e^x\left(\cos(y) + i\sin(y)\right) \qquad \text{Euler's Equation,}$$

where e is the (Euler) constant

$$e = 1 + \frac{1}{1} + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4} + \cdots$$

This polar representation of complex numbers simplifies multiplication and exponents in that

$$r_1 e^{i\alpha} \cdot r_2 e^{i\beta} = (r_1 \cdot r_2) e^{i(\alpha+\beta)} \quad \text{and} \quad \left(re^{i\theta}\right)^n = r^n e^{in\theta}.$$

Hence we can see that multiplication of complex numbers involves multiplying the moduli of the numbers and adding the angles, sometimes called an amplitwist.

We will explore these properties of the complex numbers in depth in Chapter 12.

4.7.4 Exercises

1. For each of the following complex numbers, plot the number on the complex plane and write the number in the form $a + bi$ and in the form $re^{i\theta}$. (You may leave θ as an arctangent of a number for angles that are not the standard angles.)

   a. $-14 + 14i$
   b. $\dfrac{1+i}{1-\sqrt{3}i}$
   c. $3e^{\frac{-2\pi}{3}i}$
   d. $1 + 3i$
   e. $\overline{-2 + 5i}$

2. Use the Maclaurin series representations of e^x, $\cos(x)$, and $\sin(x)$ to prove that

$$e^{iy} = \cos(y) + i\sin(y).$$

3. Use Euler's equation to prove the angle addition identities for sin and cos.

4. Use Euler's equation to prove the double angle identities for sin and cos.

5. Use Euler's equation to show De Moivre's formula

$$(\cos(x) + i\sin(x))^n = \cos(nx) + i\sin(nx)$$

   is true and use the formula to find identities for $\cos(3x)$ and $\sin(3x)$ in terms of $\cos(x)$ and $\sin(x)$.

Chapter 5

Functions

The next foundational concept of mathematics that we will study is the concept of function. Some of the early ideas of function involved describing how different quantities covaried, with the work of Oresme (1323-1382) being some of the earliest related work. Our current concept of function arose in the 17th century with the rise of analytic geometry and the foundation of calculus, with Leibniz first using the word the late 1600's [Ponte, 1992].

5.1 Definitions of Functions

Currently there are two dominant perspectives for defining a function. The first is the concept of function as a rule or procedure, as is denoted in the Common Core Standard (8.F.1).

Related Content Standards

- (8.F.1) Understand that a function is a rule that assigns to each input exactly one output. The graph of a function is the set of ordered pairs consisting of an input and the corresponding output.

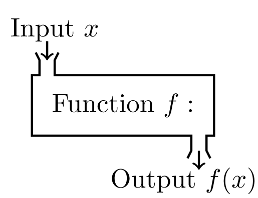


Figure 5.1: Machine analogy of a function

This definition is often represented as a box or machine with certain inputs and outputs as shown in Figure 5.1. This definition of the function, while generally very useful in understanding the use of functions, is not as mathematically sound as the generally accepted definition of function that arose in the early 1900's and appears in most theoretical mathematics textbooks. One of the main differences between the two definitions is the focus on the function being defined by three items: the domain, the codomain, and the relationship between elements of the two sets.

Definition 5.1. Let A and B be two sets, and let R be a relation from A to B. Then R is a function if for every $a \in A$, there exists a unique $b \in B$ such that $(a, b) \in R$. The set A is called the **domain** of the function and B is called the **codomain** of the function.

One should note from this definition of a function that the function is the set of ordered pairs that is often referenced as the graph of the function, rather than the algebraic or tabular representation that the input-output definition implies. Throughout the remainder of the text, we will bounce back and forth between the two views of a function, with the "input-output" or "rule" perspective often being reflected in an algebraic representation of the function and the "relation" or "ordered-pairs" perspective being reflected in a graphical representation of the function.

Related Content Standards

- (HSF.IF.1) Understand that a function from one set (called the domain) to another set (called the range) assigns to each element of the domain exactly one element of the range. If f is a function and x is an element of its domain, then $f(x)$ denotes the output of f corresponding to the input x. The graph of f is the graph of the equation $y = f(x)$.

In both of these definitions we talk about the function as a relation between two sets, A and B, and we often denote the function as $f : A \to B$. The subset of the co-domain

$$\text{Ran}(f) = \{b \in B \mid b = f(a) \text{ for some } a \in A\}$$

is called the **range** of the function. This is different than the usual definitions in the K-12 mathematics textbooks and in the Common Core State Standards where authors do not define the co-domain because it is almost always assumed to be the real numbers.

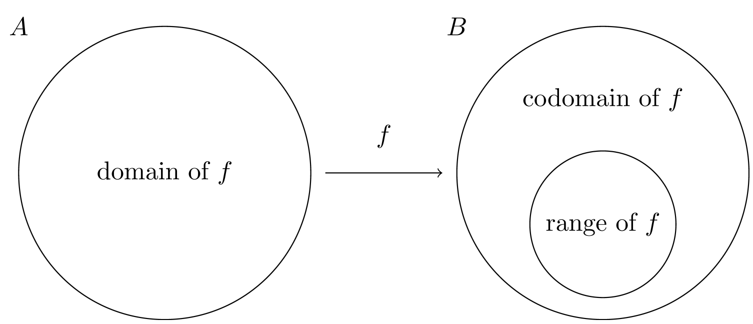


Figure 5.2: Concept sketch of $f : A \to B$

One of the key distinctions between relations that represent functions and those that do not has to do with the uniqueness of the element in the co-domain that is related to each element in the domain. For instance, with the relation from x to y given by $y = 5x$, each value of x has only one value of y related to it. However, with the relation on $[-1, 1] \times [-1, 1]$ given by $x^2 + y^2 = 1$, each value of x has two distinct values of y related to it and so y is not a function of x.

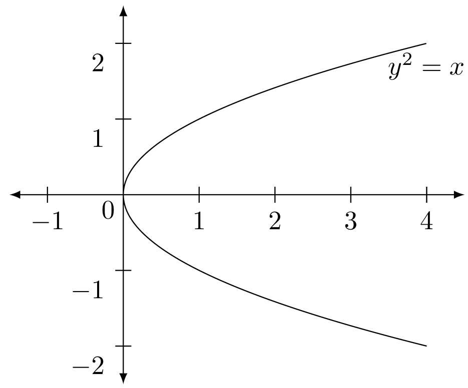
If one has the graphical representation of a relation available, one can determine if the relation represents a function by seeing if there is any value in the domain that has multiple values in the co-domain by looking for a vertical line that intersects the graph of the function at multiple points. Thus the relation on $\mathbb{R} \times \mathbb{R}$ given by $y^2 = x$ does not represent y as a function of x, but it does represent x as a function of y.

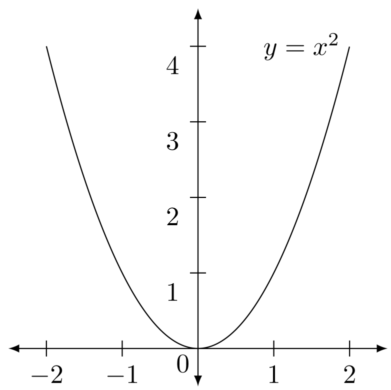
5.1.1 Function Representations

In order to understand the different representations for a function we will use the squaring function, $f : \mathbb{R} \to \mathbb{R}$, that is often represented by $f(x) = x^2$, or equivalently $x \mapsto x^2$.

If we use the sets and relation definition of the function, this function is defined to be

$$f = \left\{(x, y) \in \mathbb{R}^2 \middle| y = x^2\right\}.$$





This is often represented graphically as

If we restrict the domain and co-domain to be natural numbers, so that $f : \mathbb{N} \to \mathbb{N}$, we can represent the function in a tabular format

or by mapping elements from the domain to the co-domain

There are some times where it is easier to start describing the function by sketching how elements from the domain map to the co-domain. For example, if the function, $f : \mathbb{N} \to \mathbb{Z}$, is defined by the relationship

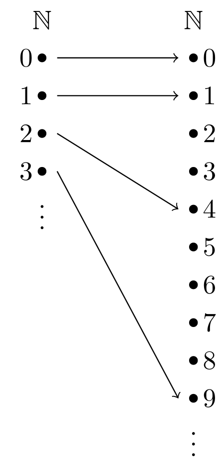
where each successive even natural number maps to the next larger positive integer and each successive odd natural number maps to the next negative integer. This function can also be represented by the graph

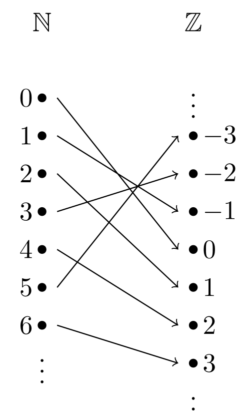
and it can be represented algebraically as

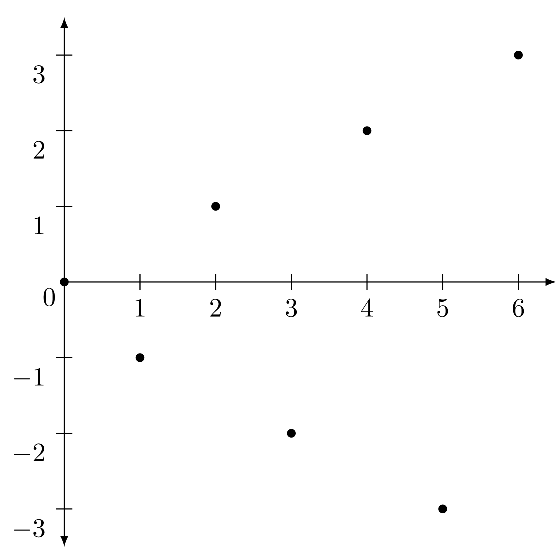
$$f(n) = \begin{cases} -\frac{n+1}{2} & \text{if } n \text{ is odd} \\ \frac{n}{2} & \text{if } n \text{ is even} \end{cases} \quad \text{or} \quad f(n) = (-1)^n \frac{2n + 1 - (-1)^n}{4}.$$

Table 5.1: The squaring function

\$x\$	\$y\$
0	0
1	1
2	4
3	9
\$\vdots\$	\$\vdots\$


ℕ
ℕ
0
1
2
3
0
1
2
3
4
5
6
7
8
9


ℕ
ℤ
0
1
2
3
4
5
6
−3
−2
−1
0
1
2
3


3
2
1
0
−1
−2
−3
1
2
3
4
5
6

5.1.2 Exercises

1. Let $A = \{a, b, c\}$ and let $B = \{1, 2\}$. List all of the possible functions from A to B.

2. Define sum : $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by $(m, n) \mapsto m + n$. Show that sum is a function.

3. Describe how the content of this section relates to the statement, "You can't take the square root of a negative number."

4. Which of the following are functions? For those that are not functions, explain why not. For those that are functions, find the range of the function.

   a. $f : \mathbb{R} \to \mathbb{R}$, with $f(x) = \frac{1}{x}$.
   b. $f = \{(x, y) \in \mathbb{R}^2 |\, x^2 + y^2 = 1\}$
   c. $f = \{(x, y) \in [-1, 1] \times [0, 1] |\, x^2 + y^2 = 1\}$
   d. $g : [0, \infty) \to \mathbb{R}$, with $g(x) = \sqrt{x}$.
   e. $h : \mathbb{N} \to \mathbb{N}$, with $h(x) = x^2$.

5. Let $f : A \to B$ be a function. Define a relation on A by

$$a_1 \sim_f a_2 \Leftrightarrow f(a_1) = f(a_2)$$

   for all $a_1, a_2 \in A$.

   a. Show that $\sim_f$ is an equivalence relation.
   b. If $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = \sin(2\pi x)$, find the equivalence class of 0 under this equivalence relation.

5.2 Injections, Surjections, and Bijections

In this section we will define and explore several different properties that some functions have that provide useful information about the function.

Definition 5.2. A function $f : A \to B$ is called an **injection**, or **one-to-one**, if no two elements of the domain map to the same element of the co-domain. Or equivalently, for elements $a_1, a_2 \in A$,

$$(f(a_1) = f(a_2)) \Rightarrow (a_1 = a_2)\,.$$

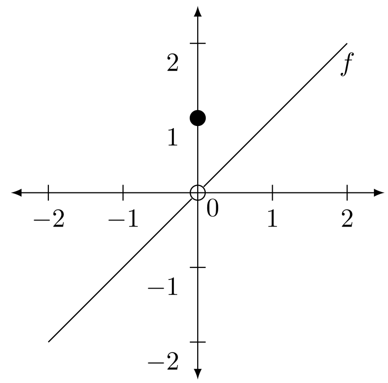
The primary benefit of a function being an injection is that for each element, b, in the range of the function there is a unique element, a, in the domain of the function such that $f(a) = b$. This provides the foundation for 'undoing' functions that we will explore further in the next section.

The simplest example of a function that is not one-to-one is given by a function from $A = \{a, b\}$ to $B = \{c\}$ represented by

$$\begin{tikzcd} A & B \\ a\bullet \arrow[dr] & \\ & \bullet c \\ b\bullet \arrow[ur] & \end{tikzcd}$$

since $f(a) = f(b)$, but $a \neq b$.

We can also see that the function $f(x) = |x|$ is not one-to-one because $f(-x) = f(x) = |x|$ for all $x \in \mathbb{R}$.



The function

$$f(x) = \begin{cases} x & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases}$$

is not one-to-one since $f(1) = f(0)$. If we sketch the graph of this function, we can see that this

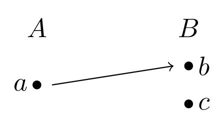
is true because if one draws a horizontal line at $y = 1$, it intersects the function at two places. This intersection at multiple places by a horizontal line is called the **horizontal line test**.

Sometimes a function cannot be graphed easily and we would like to prove that it is one-to-one. There is a fairly standard process for doing so using the one-to-one definition. In this process, we assume that a and b are in the domain of a function $f : A \to B$ such that $f(a) = f(b)$. Then using some logical arguments we prove that this implies that $a = b$. To prove that a function is not one-to-one we only need to find one case of two distinct elements in the domain mapping to the same element in the co-domain.

Definition 5.3. A function $f : A \to B$ is called a **surjection**, or **onto**, if every element of the co-domain is in the range of f. Or equivalently, if for all $b \in B$ there exists a $a \in A$ such that $f(a) = b$.

A primary consideration in a function being a surjection is to know that every element in the co-domain is 'hit' by an element in the domain. It is often more important to fully describe the range of the function, rather than if a function is a surjection, since any function is a surjection when the co-domain of the function is replaced by the range of the function in the definition of the function. For instance, $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not a surjection, but $f : \mathbb{R} \to \mathbb{R}^+$ defined by $f(x) = x^2$ is a surjection.

The basic example of a function that is not a surjection is given by a function $f : A = \{a\} \to B = \{b, c\}$ represented by



Definition 5.4. A function $f : A \to B$ is called a **bijection**, or **one-to-one and onto**, if it is both an injection and a surjection.

One of the results of a function being a bijection is that there is a perfect correspondence between elements of the domain and elements of the co-domain that we will explore further in the next two sections.

5.2.1 Exercises

1. For each of the following functions, determine if it is an injection, a surjection, and/or a bijection.

   a. $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x$.

b. $f : \mathbb{R} \to [0, \infty)$ defined by $f(x) = |x|$.
c. $f : [0, \infty) \to \mathbb{R}$ defined by $f(x) = x^2$.
d. $f : \mathbb{R} \setminus \{0\} \to \mathbb{R} \setminus \{0\}$ defined by $f(x) = \frac{1}{x}$.
e. $f : \mathbb{N} \to \mathbb{R}$ defined by

$$f(x) = \frac{x}{1 + x^2}.$$

5.3 Composition of Functions

We often want to look at how two functions work together.

Example 5.1. Suppose that oil is spilled from a ruptured tanker and spreads is a circular pattern. If the oil is leaving the tanker at a rate of 70 gallons an hour, then what is the radius of the 1 cm thick oil slick after t hours?

In order to find a solution to this word problem (note that this is not really a modelling problem) we will create some functions that we will compose together. We will start by letting g be the rate in gallons per hour that oil is leaking out of the boat.

Let $T(g) = \frac{1}{264.172} g$ be the cubic meters equivalent to g gallons.

Let $r(V) = \sqrt{\frac{100V}{\pi}}$ be the radius (in meters) of an oil slick with volume V (in cubic meters) that is 1 cm thick.

Then $r(V(T(g) \cdot t))$ gives us the radius of the oil slick after t hours.

Using this motivation, we will give a mathematical definition for the composition of functions.

Definition 5.5. Let $f : A \to B$ and $g : B \to C$ be functions. Then we define the function $(g \circ f) : A \to C$ by

$$(g \circ f)(a) = g(f(a)), \quad \text{for all } a \in A.$$

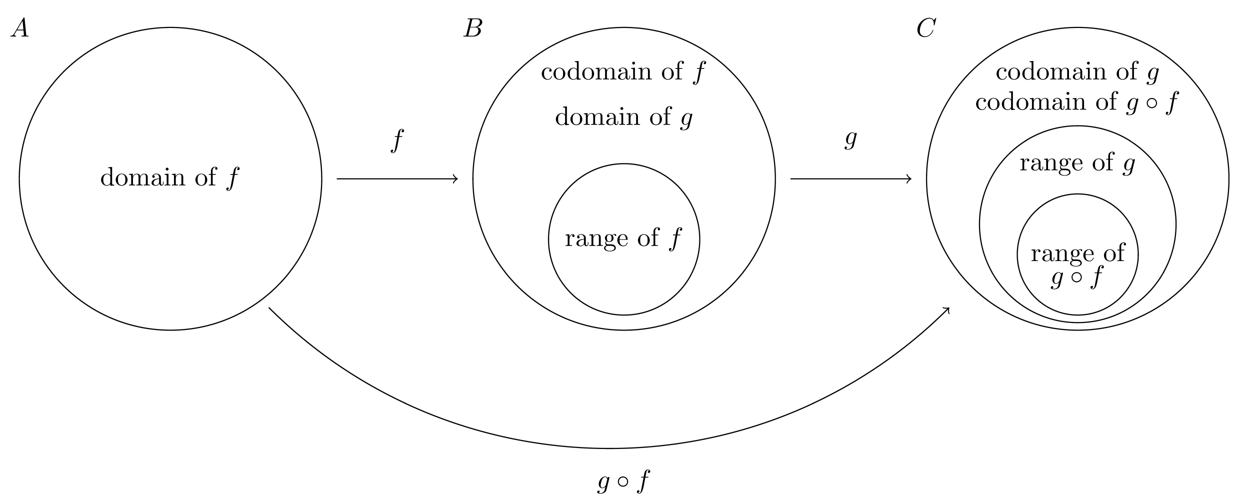


Figure 5.3: Sketch of $f : A \to B$, $g : B \to C$, and $g \circ f : A \to C$

The sketch in Figure 5.3 describes how the domains, ranges, and codomains of the three functions interact when describing function composition.

5.3.1 Inverse Functions

In addition to composing functions together, it is often useful to define a function that undoes another function. In the previous example, it is easiest to find the radius of the oil slick based on volume by first finding the volume as a function of the radius. In the case of a cylindrical cylinder, $V(r) = \pi r^2 h$ and so $r(V) = \sqrt{\frac{V}{\pi h}}$ is a function that undoes the volume. A more precise definition of these inverse functions is given below.

Definition 5.6. Let $f : A \to B$ and $g : B \to A$ be functions. If

$$(g \circ f)(a) = a \text{ for all } a \in A, \text{ and}$$
$$(f \circ g)(b) = b \text{ for all } b \in B,$$

then we say that f is **invertible**.

Note that in our example we define $f : \mathbb{R}^+ \to \mathbb{R}^+$ by $f(r) = \pi r^2 h$ to be the volume function and $g : \mathbb{R}^+ \to \mathbb{R}^+$ by $g(V) = \sqrt{\frac{V}{\pi h}}$ to be the radius function. Then for all $r \in \mathbb{R}^+$,

$$(g \circ f)(r) = g(f(r)) = g\left(\pi r^2 h\right) = \sqrt{\frac{(\pi r^2 h)}{\pi h}} = \sqrt{r^2} = r$$

and

$$(f \circ g)(V) = f(g(V)) = f\left(\sqrt{\frac{V}{\pi h}}\right) = \pi\left(\sqrt{\frac{V}{\pi h}}\right)^2 h = V.$$

If instead, we define $f : \mathbb{R} \to \mathbb{R}^+$ by $f(x) = x^2$ and $g : \mathbb{R}^+ \to \mathbb{R}$ by $g(y) = \sqrt{y}$, then

$$(f \circ g)(y) = (\sqrt{y})^2 = y, \ \forall y \in \mathbb{R}^+$$

but

$$(g \circ f)(x) = \sqrt{x^2} = |x|$$

and so f and g are not invertible.

The next theorem is useful to verify that if a function is invertible, then the inverse function is unique.

Theorem 5.1. *If $f : A \to B$ is invertible, the function $g : B \to A$ satisfying the properties*

$$(g \circ f)(a) = a \text{ for all } a \in A, \text{ and}$$
$$(f \circ g)(b) = b \text{ for all } b \in B$$

is unique and so we say that that g is the inverse function of f. We then denote this inverse function as f^{-1}.

Furthermore, we have that f^{-1} is invertible.

Proof. In order to prove the uniqueness of g we will give a proof by contradiction. So we assume that $f : A \to B$ is an invertible function and that there are function $g : B \to A$ and $\tilde{g} : B \to A$ such that

$$(g \circ f)(a) = a \text{ and } (\tilde{g} \circ f)(a) = a \text{ for all } a \in A, \text{ and}$$
$$(f \circ g)(b) = b \text{ and } (f \circ \tilde{g})(b) = b \text{ for all } b \in B.$$

Then for each $b \in B$ we have that

$$\begin{aligned}\tilde{g}(b) &= \tilde{g}\left((f \circ g)(b)\right) \text{ (since } (f \circ g)(b) = b, \text{ for all } b \in B) \\ &= \tilde{g}\left(f(g(b))\right) = (\tilde{g} \circ f)(g(b)) \text{ (by the definition of function composition)} \\ &= g(b) \text{ (since } (\tilde{g} \circ f)(a) = a, \text{ for all } a \in A).\end{aligned}$$

Since g and $\tilde{g}$ have the same output for each input, they are the same function. Therefore, the inverse function of f is unique. □

Now that we have the uniqueness of inverse functions when a function is invertible, it would be helpful to further understand the properties of a function that make it invertible without having to find the inverse. This will save time looking for an inverse that doesn't exist and help us to adjust a function to make it invertible by restricting the domain and/or co-domain of the function.

Theorem 5.2. *The function $f : A \to B$ is invertible if and only if it is a bijection.*

Proof. Since this statement is an 'if and only if' statement we must prove both implications.

- **Invertible implies bijection.** Assume that the function $f : A \to B$ is invertible. Then there exists a unique $f^{-1} : B \to A$ such that $(f \circ f^{-1})(b) = b$ for all $b \in B$ and $(f^{-1} \circ f)(a) = a$ for all $a \in A$.

  To prove that f is an injection (one-to-one), we assume that there are elements $a_1, a_2 \in A$ such that $f(a_1) = f(a_2)$ as an element of B. Since f^{-1} is a function from B to A, we know that $f^{-1}(f(a_1)) = f^{-1}(f(a_2))$. From the definitions of function composition and inverse function, this implies that

  $$a_1 = (f^{-1} \circ f)(a_1) = f^{-1}(f(a_1)) = f^{-1}(f(a_2)) = (f^{-1} \circ f)(a_2) = a_2$$

  and so $a_1 = a_2$ proving that f is an injection.

  To prove that f is a surjection (onto) we choose a generic element of the co-domain of f and prove that it is in the range of f. Let $b \in B$ be that generic element of the co-domain of f. Since

  $$b = (f \circ f^{-1})(b) = f(f^{-1}(b))$$

  we see that b is the image of f acting on $f^{-1}(b) \in A$ and so is in the range of f. Therefore, f is a bijection.

- **Bijection implies invertible.** On the other hand, if we assume that $f : A \to B$ is a bijection, we can define a function $g : B \to A$ in the following way.

  Since f is a surjection, for each $b \in B$ there is at least one element $a \in A$ such that $b = f(a)$. Since f is an injection, this element is unique and we will call the element a_b. Thus we can define $g : B \to A$ by $g(b) = a_b$ and this relation satisfies the definition of a function.

  We can also see from this definition of g that $(f \circ g)(b) = f(g(b)) = f(a_b) = b$ for all $b \in B$.

  Furthermore, for all $a \in A$, $(g \circ f)(a) = g(f(a)) = a$, thereby making f invertible.

$\square$

This means that this concept of bijections is one that is very important in terms of the composition of functions. In particular it is important to understand what properties the composition of two bijections would have. This leads to the following theorem.

Theorem 5.3. *Let $f : A \to B$ and $g : B \to C$ be two bijections. Then the function $g \circ f : A \to C$ is also a bijection.*

Proof. Let $f : A \to B$ and $g : B \to C$ be bijections.

We will first prove that $(g \circ f) : A \to C$ is an injection.

Let a_1 and a_2 be elements of A such that

$$(g \circ f)(a_1) = (g \circ f)(a_2).$$

From the definition of the composition of functions this implies that

$$g\left(f(a_1)\right) = g\left(f(a_2)\right).$$

Since g is an injection on B, we know that $f(a_1) = f(a_2)$. Since f is an injection, this implies that $a_1 = a_2$ and thus proving that $g \circ f$ is an injection.

We will now prove that $(g \circ f) : A \to C$ is a surjection.

Let c be a generic element of C. Since $g : B \to C$ is onto, there exists an element $b_c \in B$ such that $g(b_c) = c$. Since $f : A \to B$ is onto, there exists an element $a_{b_c} \in A$ such that $f(a_{b_c}) = b_c$. We then have that

$$c = g(b_c) = g\left(f\left(a_{b_c}\right)\right) = (g \circ f)\left(a_{b_c}\right)$$

and so $g \circ f$ is a surjection.

Therefore, $(g \circ f) : A \to C$ is a bijection. □

Note that in the proof that $(g \circ f) : A \to C$ is an injection that the only assumptions used are that f and g are both injections and the only sets referenced are the domains of the functions. This signifies that the property of a function being one-to-one is a property that focuses on the domain of the function and is independent of the co-domain defined for the function. Hence, one could enlarge the co-domain of an injective function to a new function that is also an injection.

Similarly, the surjective property of the composition of the functions depends only on the surjective properties of the individual functions. However, unlike the injective property, the surjective property involves a heavy reliance upon both the domain and co-domain of the function. While one could expand the domain to a larger set while maintaining the surjective nature of the function, one may not be able to reduce the size of the domain.

Since the composition of bijections is also a bijection, we know from Theorem 5.2 that the composition of a bijection is also invertible. Combining the results of these theorems we have the following theorem.

Theorem 5.4. *Let $f : A \to B$ and $g : B \to C$ be bijections. Then we have the following:*

- *f is invertible with inverse $f^{-1} : B \to A$*
- *g is invertible with inverse $g^{-1} : C \to B$*
- *$g \circ f$ is invertible with inverse $(g \circ f)^{-1} : C \to A$*
- *$(g \circ f)^{-1} = f^{-1} \circ g^{-1}$*

We can illustrate this theorem with Figure 5.4.

5.3.2 Exercises

1. Let $f : A \to B$ and $g : B \to C$ be functions. Prove that if $g \circ f : A \to C$ is a bijection that f must be an injection and g must be a surjection.
2. What adjustments can be made to the domain and/or the co-domain of $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ in order to make it an invertible function.
3. Let $f : A \to B$ and $g : B \to C$ be functions.
   a. What properties hold for $g \circ f$ if both f and g are injections?
   b. What properties hold for $g \circ f$ if both f and g are surjectionns?
   c. What properties hold for $g \circ f$ if f is an injection and g is a surjection?
   d. What properties hold for $g \circ f$ if f is a surjection and g in an injection?

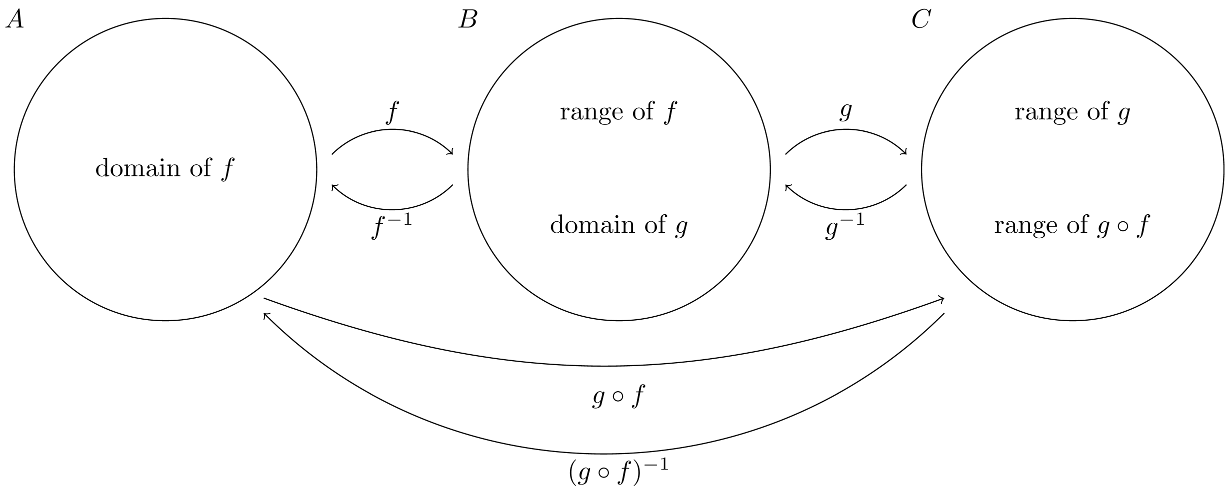


Figure 5.4: Composition of Bijections

5.4 Counting and Cardinality

The concept of cardinality is introduced in kindergarten where students are taught to have the abstract concept of a number associated to a certain number of objects such as in Figure 5.5 where we have five different representations of the same cardinality. We can give a more precise mathematical definition of cardinality using functions.

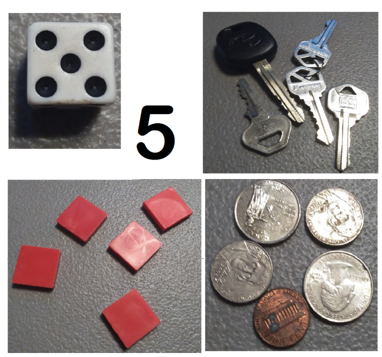


Figure 5.5: Multiple representations of the number five

Related Content Standards

- (K.CC.3) Write numbers from 0 to 20. Represent a number of objects with a written numeral $0 - 20$ with (0 representing a count of no objects).
- (K.CC.4)] Understand the relationship between numbers and quantities; connect counting to cardinality.

  a. When counting objects, say the number names in the standard order, pairing each object with one and only one number name and each number name with one and only one object.
  b. Understand that the last number name said tells the number of objects counted. The number of objects is the same regardless of their arrangement or the order in which they were counted.
  c. Understand that each successive number name refers to a quantity that is one larger.

Definition 5.7. A set A is said to have the same **cardinality** as the set B if there exists a bijection, $f : A \to B$.

This definition defines a relation on sets where A is related to B if A has the same cardinality as B.

Since the identity function, $\iota : A \to A$ defined by $\iota(a) = a$ for all $a \in A$, is a bijection we see that A has the same cardinality as itself and so the relation is reflexive.

In the previous section we proved that a bijection has a functional inverse, so if A has the same cardinality as B, there is a bijection $f : A \to B$ and a second bijection, $f^{-1} : B \to A$ showing that B has the same cardinality as A. So the cardinality relation is symmetric.

If we assume that A has the same cardinality as B and B has the same cardinality as C, there are bijections $f : A \to B$ and $g : B \to C$. In the previous section we proved that the compositions of bijections are bijections, so we know that $(g \circ f) : A \to C$ is a bijection and so A has the same cardinality as C, making the cardinality relation transitive.

These three properties of reflexive, symmetric, and transitive make the cardinality relation an equivalence relation and we define the abstract concept of a number with the equivalence class associated to that number, as defined by our set theory definition of the natural numbers.

While this might seem like a very abstract way to define counting of finite sets, it is an explanation of what is happening in the mind of a child when he is learning to count. Teachers understanding this level of abstraction can feel more empathy for the struggles of students as they learn concepts as simple as counting.

Since cardinality is an equivalence relation, for sets with a finite number of elements we can define a function

$$\# : \{\text{sets with a finite number of elements}\} \to \mathbb{N}$$

by $\#(A)$ is the cardinality of A. For sets with an infinite number of elements, we will label $\#(A)$ as the equivalence class of A using the cardinality relation.

The pigeonhole principle states that if you are putting items into containers and there are more items than containers that at least one of the containers has more than one item. Some of the earliest references to this principle is in the 1620's where multiple editions of the same book from the Jesuit university at Pont-à-Mousson referenced that there must be two men on Earth with the same number of hairs on their heads [Leurechon, 1629]. The naming of the principle after pigeonholes was based on the reference to items being placed into boxes.

Figure 5.6: Illustration of pigeonholes

In the language of functions, if we have two sets, A and B, with $\#(A) > \#(B)$, then there cannot exist an injection $f : A \to B$. In the same vein, we can say that if $\#(A) < \#(B)$, then there cannot exist a surjection $f : A \to B$. With this principle we have the following theorem.

Theorem 5.5. *Let $f : A \to B$ be a function.*

- *If f is an injection (one-to-one), then $\#(A) \leq \#(B)$.*
- *If f is a surjection (onto), then $\#(A) \geq \#(B)$.*

- *If f is a bijection, then $\#(A) = \#(B)$.*

This means that if we want to prove that two sets, A and B, have the same cardinality, we can find a bijection between them. We could also find two different functions, $f : A \to B$ and $g : A \to B$, such that one is an injection and the other is a surjection. To prove that two sets do not have the same cardinality is usually done using a proof by contradiction to show that no bijection could exist.

We explore this further using subsets of the natural numbers and an idea introduced by David Hilbert in a series of lectures for the general public in 1924 [Ewald and Sieg, 2013]. In this example we imagine a hotel with an infinite number of rooms with numbers on the doors, $1, 2, 3, 4, \ldots$, and that every room is full. When someone comes up to the registration desk, the person at the desk says that all the rooms are currently full, but we can make room for one more. The person checking in asks, "How can there be room for one more if all of the rooms are full?" The clerk says, "It's easy. We will ask each person to just gather all of their belongings and they can move to the room with the next higher number." (See Figure 5.7)

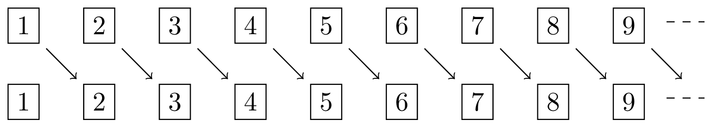


Figure 5.7: Hilbert's Hotel with one extra room.

This idea is described with function notation by the function $f : \mathbb{N}^+ \to \mathbb{N}^+$ such that $f(n) = n + 1$ as being an injective function that is not surjective. However, the function $g : \{1, 2, 3, 4, \ldots\} \to \{2, 3, 4, 5, \ldots\}$ defined by $g(n) = n + 1$ is a bijection and so the two sets have the same cardinality.

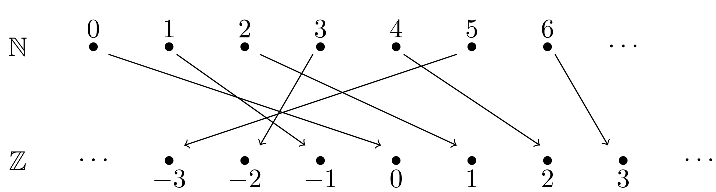
One can go a little farther in this direction by having everyone move to the room with a number that is twice their current room number. In function notation this describes a function $f : \{1, 2, 3, 4, 5, \ldots\} \to \{2, 4, 6, 8, 10, 12, \ldots\}$ such that $f(n) = 2n$, which is a bijection. So even though "half" of the numbers in the first set are removed, the two sets have the same cardinality.

This leads to the conjecture that all sets that are not of finite cardinality are the same cardinality. We will see that is not the case. We will start by dividing cardinality of infinite sets into two categories. If the set has the same cardinality as the natural numbers we will call the set **countable**, since the process of counting is a function that maps the natural numbers onto the set. If an infinite set is not countable, we will call it **uncountable**.

If a set A is countable, then we will say that $\#(A) = \aleph_0$ (called aleph-naught).

The function $f : \mathbb{N} \to 2\mathbb{N}$ such that $f(n) = 2n$ is a bijection and so both of these sets are countable.

Using the function described by



we can see that the integers are also countable.

We define $\mathbb{Q}^+$ to be the non-negative rational numbers and $f : \mathbb{N} \to \mathbb{Q}^+$ to be the function $f(n) = \frac{n}{1}$. We can see that f is an injection and so $\#(\mathbb{N}) \leq \#(\mathbb{Q}^+)$. However, if we write all of the non-negative rational expressions into a grid pattern as in Figure 5.8 we can see that we would write every possible non-negative rational expression of integers.

We can now define a function g from the natural numbers onto this set of rational expressions by mapping 0 to the expression in the first row and first column, 1 to the expression in the first row and second column, 2

$$
\begin{array}{ccccccccc}
\frac{0}{1}(0) & \frac{0}{2}(1) & \frac{0}{3}(3) & \frac{0}{4}(6) & \frac{0}{5}(10) & \frac{0}{6}(15) & \frac{0}{7}(21) & \frac{0}{8}(28) & \cdots \\
\frac{1}{1}(2) & \frac{1}{2}(4) & \frac{1}{3}(7) & \frac{1}{4}(11) & \frac{1}{5}(16) & \frac{1}{6}(22) & \frac{1}{7}(29) & \frac{1}{8}(37) & \cdots \\
\frac{2}{1}(5) & \frac{2}{2}(8) & \frac{2}{3}(12) & \frac{2}{4}(17) & \frac{2}{5}(23) & \frac{2}{6}(30) & \frac{2}{7}(38) & \frac{2}{8}(47) & \cdots \\
\frac{3}{1}(9) & \frac{3}{2}(13) & \frac{3}{3}(18) & \frac{3}{4}(24) & \frac{3}{5}(31) & \frac{3}{6}(39) & \frac{3}{7}(48) & \frac{3}{8}(58) & \cdots \\
\frac{4}{1}(14) & \frac{4}{2}(19) & \frac{4}{3}(25) & \frac{4}{4}(32) & \frac{4}{5}(40) & \frac{4}{6}(49) & \frac{4}{7}(59) & \frac{4}{8}(70) & \cdots \\
\frac{5}{1}(20) & \frac{5}{2}(26) & \frac{5}{3}(33) & \frac{5}{4}(41) & \frac{5}{5}(50) & \frac{5}{6}(60) & \frac{5}{7}(71) & \frac{5}{8}(83) & \cdots \\
\frac{6}{1}(27) & \frac{6}{2}(34) & \frac{6}{3}(42) & \frac{6}{4}(51) & \frac{6}{5}(61) & \frac{6}{6}(72) & \frac{6}{7}(84) & \frac{6}{8}(97) & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}
$$

Figure 5.8: Countability argument of non-negative rational numbers

to the expression in the second row and first column, 3 to the expression in the first row and third column, 4 to the expression in the second row and second column, 5 to the expression in the third row and first column, and so on as written in the parentheses.

We can see then that g is a surjection and so $\#(\mathbb{N}) \geq \#(\mathbb{Q}^+)$. So $\mathbb{Q}^+$ must be countable.

We can then create a very similar map from the negative integers onto the negative rational expressions and combine these two functions to create a surjection from $\mathbb{Z}$ onto $\mathbb{Q}$ and so the entire set of rational numbers is countable.

One might think that since the rational numbers are countable that all infinite sets are the same size. Cantor [1891] prove this not to be true.

Theorem 5.6. *The open interval $(0,1)$ in the real numbers is not countable.*

Proof. We will follow the general flow of a proof by Georg Cantor (1891). If we assume that the interval $(0,1)$ is countable, then there is a bijection $f : \mathbb{N} \to (0,1)$. For each natural number, n, we will write $f(n)$ in a decimal representation as

$$f(n) = 0.a_{1,n}a_{2,n}a_{3,n}\ldots = \sum_{i=1}^{\infty} a_{i,n} \cdot 10^{-i},$$

where each of the $a_{i,n} \in \{0,1,2,3,4,5,6,7,8,9\}$, and in order to avoid two different decimal expansions representing the same real number we will assume that the $a_{i,n}$ are not eventually all 9.

For each $j \in \{1,2,3,4,\ldots\}$, let b_j be chosen from $\{1,2,3,4,5,6,7,8\}$ such that $b_j \neq a_{j,j}$. Then $b = \sum_{i=1}^{\infty} a_{i,n} \cdot 10^{-i}$ is a real number in $(0,1)$ such that $b \neq f(n)$ for any natural number n. This means that f was not a surjection and so no such bijection exits. $\square$

Now that we know that there are sets that are not countable, let's explore some other uncountable sets.

Let $f : (0,1) \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} \frac{2x-1}{1-x}, & \text{if } \frac{1}{2} \leq x < 1 \\ \frac{2x-1}{x}, & \text{if } 0 < x \leq \frac{1}{2} \end{cases}.$$

If we choose a generic real number and call it b, then if $b > 0$, we have that

$$f\left(\frac{1+b}{2+b}\right) = \frac{2\left(\frac{1+b}{2+b}\right) - 1}{1 - \left(\frac{1+b}{2+b}\right)} = \frac{(2+2b)-(2+b)}{(2+b)-(1+b)} = b,$$

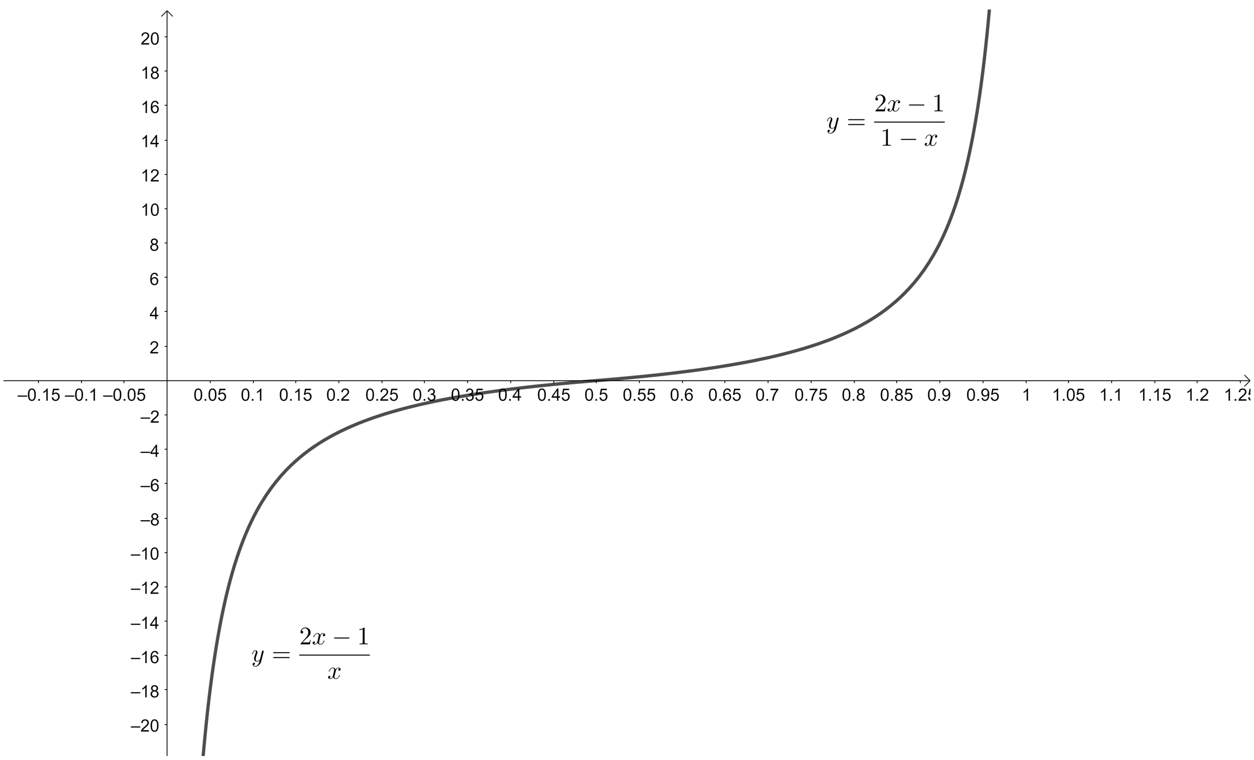


Figure 5.9: Example of an uncountable bijection ([Corresponding Geogebra Page](https://www.geogebra.org/classic/jyun3jnu))

and if $b < 0$,

$$f\left(\frac{1}{2-b}\right) = \frac{2\left(\frac{1}{2-b}\right)-1}{\left(\frac{1}{2-b}\right)} = \frac{2-(2-b)}{1} = b.$$

Since $f\left(\frac{1}{2}\right) = 0$, we have that f is onto and so $\#((0,1)) \geq \#(\mathbb{R})$. Since $(0,1) \subseteq \mathbb{R}$, we have the $\#((0,1)) \leq \mathbb{R}$. With both inequalities being true, we have that the two sets have the same cardinality.

In fact, any subset of the real numbers is either finite, countable, or the same cardinality as $\mathbb{R}$. Cantor called the cardinality of $\mathbb{R}$ the cardinality of the continuum and also proved that it is the same cardinality as the power set of the natural numbers ($\#(\mathbb{R}) = \#(\mathcal{P}(\mathbb{N}))$.

5.4.1 Exercises

1. Prove that the following pairs of sets have the same cardinality.

   a. $3\mathbb{Z} = \{n \in \mathbb{Z} | n = 3k \text{ for some } k \in \mathbb{Z}\}$ and $5\mathbb{Z} = \{n \in \mathbb{Z} | n = 5k \text{ for some } k \in \mathbb{Z}\}$
   b. $[a, b]$ and $[c, d]$, where a, b, c, and d are real numbers such that $a < b$ and $c < d$.
   c. A and $A \cup B$, where A and B are countable sets.

2. Prove that if A is a set of finite cardinality, the cardinality of the power set of A is $2^{\#(A)}$.

3. Look up a proof that the algebraic numbers are countable, and compare that to the proof that the rational numbers are countable.

4. Define two sets (not necessarily with finite cardinality) to be equivalent if they have the same cardinality. For a set A, let $\mathcal{P}(A)$ be the set of all subsets of A. Prove that A is not equivalent to $P(A)$. [Hint:

Suppose $f : A \to P(A)$ and define

$$C = \{x | x \in A \text{ and } x \notin f(x)\}.$$

Show $C \notin \operatorname{im}(f)$.]

Part II

ALGEBRA

Chapter 6

Groups, Rings, and Fields

Abstract (or modern) algebra studies basic mathematical structures that appear in many settings . In terms of groups, we study mathematical sets and a single operation (such as addition or multiplication). The study of how two operations interact is contained in studying rings and fields. In this chapter we intersperse the axioms and definitions of abstract algebra with the concrete examples that appear in the K-12 curriculum. The goal of the study of these basic groups, rings, and fields is to better understand the connections between different content in the K-12 curriculum with the goal of improving mathematics teaching by helping students learn new material by making connections with material they have previously learned.

6.1 Group Theory

Definition 6.1 (Group)**.** A non-empty set G, together with a binary operation, $*$, is called a group if it satisfies the following conditions:

- $a * b \in G$, $\forall a, b \in G$ (Closure)
- $(a * b) * c = a * (b * c), \forall a, b, c \in G$ (Associative)
- There exists an element $e \in G$ such that for all $a \in G$, $e * a = a * e = a$ (Identity)
- For each $a \in G$, there exists an element $b \in G$ such that $a * b = b * a = e$ (Inverse)

This concept of a group is a generalization of most of the properties that we noticed in Chapter 4 as we constructed the various number systems.

Example 6.1. Some basic examples of groups that appear in the K-12 curriculum include:

1. $(\mathbb{Z}, +)$, the set of integers under addition
2. $(2\mathbb{Z}, +)$, the set of even integers under addition
3. $(\mathbb{Q}, +)$, the set of rational numbers under addition
4. $(\mathbb{R}, +)$, the set of real numbers under addition
5. $(\mathbb{Q} \setminus \{0\}, \cdot)$, the set of non-zero rational numbers under multiplication
6. $(\mathbb{R}^+, \cdot)$, the set of positive real numbers under multiplication
7. $(\mathbb{R}[x], +)$, the set of polynomials with real coefficients under addition

8. The set of one-to-one and onto functions, $f : A \to A$ for some set A, under function composition.
9. $\left(M_{(m,n)}(\mathbb{R}), +\right)$, the set of $m \times n$ matrices under addition.
10. $GL_n(\mathbb{R})$, the set of $n \times n$ invertible matrices under multiplication.
11. $SL(n, \mathbb{R})$, the set of $n \times n$ invertible matrices, with determinant of 1, under multiplication.

If we extend our examples to the calculus curriculum we can include the differentiable (or integrable) functions with the binary operation of addition. For the group of differentiable functions under addition, the identity is the function $f_0 : \mathbb{R} \to \mathbb{R}$ defined by $f_0(x) = 0$ for all x. If $g : \mathbb{R} \to \mathbb{R}$ is differentiable, then its inverse in this group would be defined by $h(x) = -g(x)$.

The cardinality of each of the sets for the groups in Example 6.1 is at least countable. We can also have groups with a finite number of elements as shown in the following example.

Example 6.2. Let $\mathbb{Z}_2 = \{0, 1\}$ with addition defined by the following table (often referred to as a Cayley table)

+	0	1
0	0	1
1	1	0

Then the identity is 0 and each element is its own inverse.

Note that the complete definition of a group involves both the set G and the binary operation $*$ and so the group is officially labeled as $(G, *)$. However, the binary operation is often inferred from the set itself. So we often just refer to the group as G with the operation implied.

When showing that a set and operation form a group, the most challenging property to show directly is often the associative property. Since almost all of the groups that we will encounter are derived from the elements and operations in the number systems constructed in Chapter 4, we will gain the associative property from those number systems.

6.1.1 Abelian Groups

In most of the above mentioned groups, the order of the operation does not matter. However, for groups such at $GL_n(\mathbb{R})$, the order of the elements with the operation does matter.

Definition 6.2 (Abelian Groups)**.** If a group $(G, *)$ also satisfies the property $a * b = b * a$, $\forall a, b \in G$ then $(G, *)$ is called **abelian**.

Example 6.3. The non-abelian group whose set has the smallest cardinality is the dihedral group of order 6 defined by the Cayley table:

$*$	e	a	b	c	d	f
e	e	a	b	c	d	f
a	a	e	d	f	b	c
b	b	f	e	d	c	a
c	c	d	f	e	a	b
d	d	c	a	b	f	e
f	f	b	c	a	e	d

We can verify that this group is not abelian since $a * b = d$ and $b * a = f$.

It will be important to notice whether or not a group is abelian, as abelian groups have a simpler structure than those for which the operation does not commute.

6.1.2 Uniqueness of Identities and Inverses

The existence of identity and inverse elements does not guarantee uniqueness, a priori. However, the following two theorems to verify the uniqueness of the identity and inverse elements within a group.

Theorem 6.1 (Uniqueness of the Identity)**.** *For a group, $(G, *)$, the identity is unique.*

Proof. Assume that there are two distinct identity elements that we will call e and $\tilde{e}$. Then the following is true:

$$\tilde{e} = e * \tilde{e} = e$$

since both e and $\tilde{e}$ are identities. Thus the identities are not distinct and so the identity is unique. □

Theorem 6.2 (Uniqueness of the Inverse)**.** *For a group, $(G, *)$, for each $a \in G$, the inverse of a is unique and so we call the inverse a^{-1}.*

Proof. Let $a \in G$ and assume that there are distinct elements $b, c \in G$ that are inverses of a. Then

$$b = b * (e) = b * (a * c) = (b * a) * c = (e) * c = c.$$

Therefore, the inverse of a is unique. □

6.1.3 Homomorphisms

Now that we have a better understanding of the definitions and structures of groups we turn our attention to looking at when two groups are really the 'same' group but viewed from different perspectives. To do that we must create a way to determine equivalence classes of groups that are based on a relationship between the two sets and how they interact with their binary operations.

Definition 6.3. A function, ϕ, from a group $(G, *)$ to a group $(G', \cdot)$ is called a **homomorphism** if

$$\phi(a * b) = \phi(a) \cdot \phi(b)$$

for every $a, b \in G$.

In other words, a homomorphism is a function between groups that maintains the structure of the binary operation. Notice that this restricts the types of functions quite a bit since a simple function like $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x + 1$ is not a homomorphism.

It will be helpful to understand this idea of maintaining the structure of the binary operation with a few examples of functions, some of which are homomorphisms and some are not.

Example 6.4. Let $\phi : (\mathbb{Z}, +) \to (\mathbb{Z}, +)$ be defined by $\phi(n) = 2n$. In order to verify that ϕ is a homomorphism, we see that for any integers m, n that

$$\phi(m + n) = 2(m + n) = 2m + 2n = \phi(m) + \phi(n).$$

So ϕ is a homomorphism.

Example 6.5. Let $f : (\mathbb{R}, +) \to (\mathbb{R}^+, \cdot)$ be defined by $f(x) = 2^x$. Then for real numbers x and y we have that

$$f(x + y) = 2^{(x+y)} = 2^x \cdot 2^y = f(x) \cdot f(y)$$

making f a homomorphism.

Example 6.6. Let $g : (\mathbb{R}, +) \to (\mathbb{R}, +)$ be defined by $g(x) = 2x + 3$. Then for real numbers x and y,

$$g(x + y) = 2 \cdot (x + y) + 3 = 2x + 2y + 3.$$

However,

$$g(x) + g(y) = (2x + 3) + (2y + 3) = 2x + 2y + 6$$

and so g is not a homomorphism.

For each of the examples above that generated a group homomorphism, notice that each of these homomorphisms mapped the identity element to the identity element and the mapped inverses to inverses. We see in the following two theorems that this is true for all homomorphisms.

Theorem 6.3. *Let $f : (G, *) \to (G', \cdot)$ be a group homomorphism. If e and e' are the identity elements of $(G, *)$ and $(G', \cdot)$, respectively, then $f(e) = e'$.*

Proof. Let a be any element of G and let e be the identity in G. Then

$$f(a) = f(a * e) = f(a) \cdot f(e)$$

and so

$$e' = f(a)^{-1} \cdot f(a) = f(a)^{-1} \cdot (f(a) \cdot f(e)) = (f(a)^{-1} \cdot f(a)) \cdot f(e) = e' \cdot f(e) = f(e).$$

□

Theorem 6.4. *Let $f : (G, *) \to (G', \cdot)$ be a group homomorphism. Then for each $a \in G$, $f(a)^{-1} = f\left(a^{-1}\right)$.*

Proof. Since $f(a) \cdot f\left(a^{-1}\right) = f\left(a * a^{-1}\right) = f(e) = e'$ and since the inverse of an element is unique, we have that $f(a)^{-1} = f\left(a^{-1}\right)$. □

Just like with the functions in Chapter 5, we often find it useful to compose two homomorphisms. We see in the following theorem that such a composition is also a homomorphism.

Theorem 6.5. *Let $f : (A, *) \to (B, \times)$ and $g : (B, \times) \to (C, \cdot)$ be two group homomorphisms. Then*

$$(g \circ f) : (A, *) \to (C, \cdot)$$

is a group homomorphism.

Proof. The proof of the theorem follows directly from the definition of f and g being group homomorphisms, with the difficulty lying in remembering the appropriate operation at each step in the process.

Let a_1 and a_2 be generic elements of A. Then

$$\begin{aligned}(g \circ f)(a_1 * a_2) &= g\left(f(a_1 * a_2)\right) = g\left(f(a_1) \times f(a_2)\right) \\ &= g(f(a_1)) \cdot g(f(a_2)) = (g \circ f)(a_1) \cdot (g \circ f)(a_2)\end{aligned}$$

and so $g \circ f$ is a homomorphism. □

6.1.4 Isomorphisms

When we combine the concept of a homomorphism with the idea of a bijection, we obtain a new type of function on groups.

Definition 6.4. A homomorphism between groups is called an **isomorphism** if it is one-to-one and onto.

Since isomorphisms are bijections, they satisfy all of the same properties of bijections, including having an inverse. We also can deduce that an isomorphism can only exist between groups whose sets have the same cardinality. So there cannot exist an isomorphism from $(\mathbb{Z}_2, +)$ in Example 6.2 to the dihedral group of order 6 in Example 6.3.

Just as a bijections create an equivalence relation in terms of cardinality, we will see that isomorphisms create an equivalence relation in terms of the group structure. In order to develop this equivalency it is helpful to understand the properties of the inverse function related to the group structures. As such, in the next theorem we prove that the inverse of an isomorphism is also an isomorphism.

Theorem 6.6. *Let $f : (G, *) \to (G', \cdot)$ be an isomorphism. Then $f^{-1} : (G', \cdot) \to (G, *)$ is also an isomorphism.*

Proof. Since f is a bijection, we know from Chapter 5 that f^{-1} is also a bijection. So in order to prove that f^{-1} is an isomorphism it suffices to prove that f^{-1} is a homomorphism.

If we let b_1 and b_2 be two generic elements of G'. Then since f is a bijection, there exist unique a_1 and a_2 in G such that $b_1 = f(a_1)$ and $b_2 = f(a_2)$. Then

$$\begin{aligned} f^{-1}(b_1 \cdot b_2) &= f^{-1}\left(f(a_1) \cdot f(a_2)\right) \\ &= f^{-1}\left(f(a_1 * a_2)\right) \quad \text{(since } f \text{ is a homomorphism)} \\ &= (f^{-1} \circ f)(a_1 * a_2) = a_1 * a_2 \\ &= f^{-1}(b_1) * f^{-1}(b_2) \end{aligned}$$

and so f^{-1} is a homomorphism, and since it is a bijection it is an isomorphism. □

Theorem 6.7. *We say that a group $(G, *)$ is isomorphic to a group $(G', \cdot)$ if there is an isomorphism from $(G, *)$ to $(G', \cdot)$. This relation of $(G, *)$ isomorphic to $(G', \cdot)$ is an equivalence relation.*

Proof. Recall that in order to prove that a relation is an equivalence relation we have to prove that the relation is reflexive, symmetric, and transitive.

- **Reflexive.** Let $(A, *)$ be a group and define $\phi : (A, *) \to (A, *)$ by $\phi(a) = a$ for all $a \in A$. Since ϕ is the identity function on $(A, *)$ we see that ϕ is an isomorphism and so $(A, *)$ is isomorphic to itself.
- **Symmetric.** Let $(A, *)$ and $(B, +)$ be groups such that $(A, *)$ is isomorphic to $(B, +)$. Let $f : (A, *) \to (B, +)$ be such an isomorphism. From the previous theorem we know that $f^{-1} : (B, +) \to (A, *)$ is also an isomorphism. Therefore, $(B, +)$ is isomorphic to $(A, *)$.
- **Transitive.** The proof that isomorphisms are transitive follows from Theorem 5.3 that the composition of two bijections is a bijection and Theorem 6.5 that the composition of two homomorphisms is a homomorphism. Therefore, the composition of two isomorphisms is again an isomorphism. Therefore, group isomorphism creates a transitive relation.

□

The concept of isomorphism in algebra is similar to the concept of congruence in geometry where two triangles being congruent means that they share all of the related properties of triangles. Two groups are considered the "same" group if there is an isomorphism between them and are distinct only in the perspective in which one is viewing the groups. Such groups will be referred to as isomorphic groups. If G and G' are isomorphic groups, then we will denote this by $G \cong G'$.

6.1.5 Finite Groups

To better understand this relationship between isomorphic groups, let's look at groups with two elements. If $(G, *)$ is a group with two elements, the identity e, and an element $a \neq e$. We know that in order for $(G, *)$ to be a group, we know that each element must have a unique inverse. Since $a * e = a$, we know that e is not the inverse of a. So the inverse of a must be a and we have the Cayley table

$*$	e	a
e	e	a
a	a	e

We can then see that the function $\phi : (\mathbb{Z}_2, +) \to (G, *)$ defined by $\phi(0) = e$ and $\phi(1) = a$, that ϕ is an isomorphism and so every group with two elements is isomorphic to $(\mathbb{Z}_2, +)$.

In order to study groups with three elements, similarly to how we defined $(\mathbb{Z}_2, +)$ in Example 6.2, we can define $(\mathbb{Z}_3, +)$ as the set $\{0, 1, 2\}$ with the operation defined by the Cayley table,

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

If we let $(G, *)$ be another group with three distinct elements, $\{e, a, b\}$, with identity, e, we know that $e * a = a$, $a * e = a$, $e * b = b$, and $b * e = b$ giving the first row and first column of the Cayley table,

$*$	e	a	b
e	e	a	b
a	a		
b	b		

Since the group is closed under the operation, $a * b \in G$. If $a * b = a$, since a has an inverse, we have $b = a^{-1} * (a * b) = a^{-1} * a = e$. Similarly, if $a * b = b$, we would conclude that $a = e$. Since the identity is unique, neither of these could be true. So $a * b = e$ and $b * a = e$, and we have filled in another two positions in the Cayley table.

$*$	e	a	b
e	e	a	b
a	a		e
b	b	e	

We now need to determine $a * a$ and $b * b$. Since inverses are unique, we know that $a * a \neq e$ and $b * b \neq e$. If $a * a = a$, we have that $a = a * e = a * a * b = a * b = e$, which is not possible. So $a * a = b$, and similarly, $b * b = a$. So we see that $(G, *)$ has the Cayley table

$*$	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

We can now define a function $\phi : (\mathbb{Z}_3, +) \to (G, *)$ by the correspondence $0 \mapsto e$, $1 \mapsto a$, and $2 \mapsto b$, and one can verify that such a function is an isomorphism. So we can now conclude that all groups with three elements are isomorphic to $(\mathbb{Z}_3, +)$. One cannot prove the same type of result for groups with four elements, since $(\mathbb{Z}_4, +)$ is not isomorphic to $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$.

$(\mathbb{Z}_4, +)$

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$

+	(0,0)	(0,1)	(1,0)	(1,1)
(0,0)	(0,0)	(0,1)	(1,0)	(1,1)
(0,1)	(0,1)	(0,0)	(1,1)	(1,0)
(1,0)	(1,0)	(1,1)	(0,0)	(0,1)
(1,1)	(1,1)	(1,0)	(0,1)	(0,0)

The classification of the groups with a finite number of elements was a significant endeavor in throughout the twentieth century, with a complete classification verified in the early twenty-first century using computers to check the validity of the proof.

6.1.6 Exercises

1. For each of the groups listed in Example 6.1:

   a. What is the identity element?

   b. For a generic element a, what would a^{-1} look like?

   c. Is the group abelian?

   d. In what ways are the groups related to one another? (Think about subsets or homomorphisms)

2. Find a homomorphism from $(\mathbb{Z}, +)$ to $(5\mathbb{Z}, +)$.

3. Are $(\mathbb{Z}, +)$ and $(\mathbb{R}, +)$ isomorphic groups? Why, or why not?

4. Let $(G, *)$ and $(H, *)$ be two groups such that $H \subseteq G$ and the operation is the same for both groups.

   a. What are some possible homomorphisms $\phi : (H, *) \to (G, *)$?

   b. What are some possible homomorphisms $\psi : (G, *) \to (H, *)$?

5. Let $\phi : (G, *) \to (H, \circ)$ be a group homomorphism.

   a. Define
   $$\ker(\phi) = \{g \in G \mid \phi(g) = e_H\}$$
   and prove that $\ker(\phi)$ is a group.

   b. Define
   $$\operatorname{ran}(\phi) = \{h \in H \mid h = \phi(g) \text{ for some } g \in G\}$$
   and prove that $\operatorname{ran}(\phi)$ is a group.

6.2 Relationship Between Linear and Exponential

In this section we will explore exponential and linear functions through the perspective of isomorphisms. Let a be a positive real number. Then we define the exponential function with base a as the function

$$\exp_a : (\mathbb{R}, +) \to (\mathbb{R}^+, \cdot)$$

with $\exp_a(x) = a^x$, where a^x is defined in Section 4.6.

Using the properties of exponents in Theorem 4.16 we have that $\exp_a$ is an increasing function (i.e. $x < y \Rightarrow \exp_a(x) < \exp_a(y)$). This implies that $\exp_a$ is an injection. Using properties of limits from analysis one can prove that $\exp_a$ is a surjection. From Section 4.6 we have that

$$\exp_a(x + y) = a^{x+y} = a^x \cdot a^y = \exp_a(x) \cdot \exp_a(y)$$

and so $\exp_a$ is an isomorphism.

We will now explore the isomorphism between $(\mathbb{R}, +)$ and $(\mathbb{R}^+, \cdot)$.

Property in $(\mathbb{R}, +)$	Property in $(\mathbb{R}^+, \cdot)$
$m + m = 2 \cdot m$	$a \cdot a = a^2$
$mk = m + m + m + \cdots + m$, k times	$a^k = a \cdot a \cdot a \cdots a$, k times
$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$	$(b \cdot c)^a = (b^a) \cdot (c^a)$
$0 + m = m + 0 = m$ (additive identity)	$1 \cdot a = a \cdot 1 = a$ (multiplicative identity)
$m + (-m) = 0$ (additive inverse)	$a \cdot \frac{1}{a} = 1$ (multiplicative inverse)

One can see that repeated addition in $(\mathbb{R}, +)$ corresponds to repeated multiplication in $(\mathbb{R}^+, \cdot)$. One can also see that

$$\exp_a(m \cdot x + b) = \exp_a(b) \cdot (\exp_a(m))^x$$

and so we can see that linear expressions in $(\mathbb{R}, +)$ correspond to exponential expressions in $(\mathbb{R}^+, \cdot)$. An implication of this is that the teaching of linear and exponential functions should be intertwined. At a minimum, teachers should use their horizon content knowledge when initially teaching students about exponential functions to directly related the properties of the functions to those of linear functions.

Related Content Standards

- (HSF.LE.1) Distinguish between situations that can be modeled with linear functions and with exponential functions.
  a. Prove that linear functions grow by equal differences over equal intervals, and that exponential functions grow by equal factors over equal intervals.
  b. Recognize situations in which one quantity changes at a constant rate per unit interval relative to another.
  c. Recognize situations in which a quantity grows or decays by a constant percent rate per unit interval relative to another.

In order to compare linear and exponential functions, we will study their properties side-by-side. Let

$$f(x) = mx + b \quad \text{and} \quad g(x) = b \cdot m^x$$

be linear and exponential functions, respectively. The graphs of the functions are then given by the set of points (x, y) such that

$$y = mx + b \quad \text{and} \quad y = b \cdot m^x.$$

Setting $x = 0$ in each equation implies that $(0, b)$ is the y-intercept for both functions. Exploring the relationship of $f(x)$ to $f(x + 1)$ and $g(x)$ to $g(x + 1)$ we see that

$$f(x + 1) - f(x) = m \quad \text{and} \quad \frac{g(x + 1)}{g(x)} = m$$

implying that in both cases the parameter m represents a constant rate of change. For the linear function m represents an additive rate of change, while in the exponential function m represents a multiplicative rate of change.

Related Content Standards

- (HSF.LE.5) Interpret the parameters in a linear or exponential function in terms of a context.

For the linear function, the function is increasing if, and only if, m is greater than the additive identity. With the exponential function, the values of m must be positive and the function is increasing if, and only if, m is greater than the multiplicative identity.

If we know that two points, (x_1, y_1) and (x_2, y_2), are on the graph of the functions, then we know that

$$\begin{aligned} y_2 &= m \cdot x_2 + b \\ y_1 &= m \cdot x_1 + b \end{aligned} \quad \text{or} \quad \begin{aligned} y_2 &= b \cdot m^{x_2} \\ y_1 &= b \cdot m^{x_1}. \end{aligned}$$

Using subtraction for the linear function and division for the exponential function we find that

$$(y_2 - y_1) = m(x_2 - x_1) \quad \text{or} \quad \frac{y_2}{y_1} = m^{(x_2 - x_1)}$$

and so

$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{or} \quad m = \left(\frac{y_2}{y_1}\right)^{\frac{1}{x_2 - x_1}}.$$

After determining the rate of change for the function one can now include that information in a point-'rate of change' form of the function as

$$y = m(x - x_1) + y_1 \quad \text{or} \quad y = y_1 \cdot m^{(x - x_1)}.$$

6.2.1 Exercises

1. Find the linear and exponential functions that pass through the points $(-3, -2)$ and $(2, 1)$.

6.3 Rings and Fields

In this section we are going to first explore several different examples of sets that have a standard pair of two different binary operations, addition and multiplication. We will create categorizations of these sets based upon their properties. Some properties to consider are associative, commutative, distributive, identities, and inverses.

Example 6.7. Consider the following sets and operations.

- $\mathbb{N}$ with the usual addition and multiplication.
- $\mathbb{Z}$ with the usual addition and multiplication.
- $\mathbb{Q}$ with the usual addition and multiplication.
- $\mathbb{R}$ with the usual addition and multiplication.
- $\mathbb{C}$ with the usual addition and multiplication.
- $2\mathbb{Z}$ (the even integers) with the usual addition and multiplication.

For each of these sets with operations, determine which of the following properties hold.

1. Additive identity
2. Additive inverse
3. Additive commutativity

4. Multiplicative identity
5. Multiplicative inverse
6. Multiplicative commutativity
7. Distribution properties of multiplication over addition.

6.3.1 Finite Sets with Two Operations

While each of the sets listed previously have infinite cardinality, a large portion of abstract algebra studies finite sets, combined with appropriate operations. $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ with addition and multiplication defined by the remainder of the sum and product when divided by n. We provide some examples of these types of sets and operations below. For these finite sets, it is often easiest to just provide a table of the operations, rather than an algebraic definition. These tables are called Cayley tables, first presented in a work by the British mathematician, Arthur Cayley [1854].

Example 6.8. Let $\mathbb{Z}_2 = \{0, 1\}$ with the operations given in the following Cayley tables.

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Example 6.9. Let $\mathbb{Z}_3 = \{0, 1, 2\}$ with the operations given in the following Cayley tables.

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Example 6.10. Let $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ with the operations given in the following Cayley tables.

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Example 6.11. Let $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$ with the operations given in the following Cayley tables.

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

These different sets (and operations) satisfy various properties depending upon the number of elements in the set. It is a valuable exercise to explore how the these are connected.

6.3.2 Matrices

We let $GL(n, \mathbb{R})$ be the set of $n \times n$ matrices with coefficients from the real numbers with the usual matrix addition and multiplication. We can see that this set has an additive identity, 0, and multiplicative identity,

I.

$$0 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \quad \text{and} \quad I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

It is also satisfies most of the properties of our number systems that we have explored, including the addition is commutative. However, we can discover that multiplication is not commutative, not every element has a multiplicative inverse, and we can have the product of two non-zero elements be zero. For example, if we look at $GL(2,\mathbb{R})$ and the matrices

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Then $A^2 = 0$, $B^2 = 0$,

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad BA = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Sometimes we combine ideas of the finite sets with the additional structure of matrices. For example, we let $GL(2,\mathbb{Z}_2)$ be the two by two matrices with coefficients from $\mathbb{Z}_2$, with the usual operations of matrix addition and multiplication.

Related Content Standards

- (HSN.VM.8) Add, subtract, and multiply matrices of appropriate dimensions.
- (HSN.VM.9) Understand that, unlike multiplication of numbers, matrix multiplication for square matrices is not a commutative operation, but still satisfies the associative and distributive properties.
- (HSN.VM.10) Understand that the zero and identity matrices play a role in matrix addition and multiplication similar to the role of 0 and 1 in the real numbers. The determinant of a square matrix is nonzero if and only if the matrix has a multiplicative inverse.

6.3.3 Rings

Now that we have several examples of sets with two binary operations defined on them, we see that there are certain properties of these binary operations that are very useful for understanding the structure of the set.

Definition 6.5 (defrings)**.** A **ring** $\langle R, +, \cdot \rangle$ is a set R together with two binary operations $+$ and $\cdot$, which we call addition and multiplication, defined on R such that the following are satisfied:

1. $(R, +)$ is an abelian group.
   - The binary operation $+$ is associative.
   - There is an element 0 in R such that $0 + a = a + 0 = a$ for all $a \in R$. (This element is called the additive identity of R.)
   - For each a in R there is an element $-a$ in R such that $(-a) + a = a + (-a) = 0$. (This element is called the additive inverse of a in R.)
   - For each a and b in R, $a + b = b + a$. (This means that the operation is commutative and so the group is abelian.)
2. The binary operation $\cdot$ is associative.
3. For all $a, b, c \in R$, the left distribution law, $a(b + c) = (ab) + (ac)$, and the right distribution law, $(a + b)c = (ac) + (bc)$, hold.

We can see from this definition that the integers, $\langle \mathbb{Z}, +, \cdot \rangle$, form a ring but the natural numbers, $\langle \mathbb{N}, +, \cdot \rangle$, do not because they do not have additive inverses. Similarly, during the constructions of the rational numbers, $\langle \mathbb{Q}, +, \cdot \rangle$; the real numbers, $\langle \mathbb{R}, +, \cdot \rangle$; and the complex numbers, $\langle \mathbb{C}, +, \cdot \rangle$, we proved that these number systems satisfied all of the ring properties.

We can also see that the examples above of $\langle \mathbb{Z}_n, +, \cdot \rangle$ and $\langle GL(n, \mathbb{R}), +, \cdot \rangle$ are also rings. So in many ways, all of these sets and operations have some very similar properties. However, there are some properties that some of these rings have, that others do not. If a ring has certain additional properties, we will add some labels to these sets.

Definition 6.6 (Further definitions)**.** We classify rings based on various properties of the ring.

1. A ring in which the multiplication is commutative is called a **commutative ring**.
2. A ring with a multiplicative identity is called a **ring with unity**.
3. If every non-zero element of a ring has a multiplicative inverse, then the ring is called a **division ring**.
4. A **field** is a commutative division ring.

Related Content Standards

- (7.NS2) Apply and extend previous understandings of multiplication and division and of fractions to multiply and divide rational numbers.
  a. Understand that multiplication is extended from fractions to rational numbers by requiring that operations continue to satisfy the properties of operations, particularly the distributive property, leading to products such as $(-1)(-1) = 1$ and the rules for multiplying signed numbers. Interpret products of rational numbers by describing real-world contexts.

6.3.4 Identities and Inverses

Now that we have created several categories and labels for the sets and operations, let's explore some of the consequences of the properties on the operations. As we do so, keep in mind which properties are used to prove different results. That way you can determine which properties are required of various sets and operations in order to get those results later on.

Theorem 6.8. *The additive identity for a ring is unique.*

This theorem is a direct consequence of the set with addition forming a group, whose identity is unique. Similarly, we have that the multiplicative identity is unique (if it exists). Since the proof is so similar to that of the uniqueness of the additive identity, the proof is left to the reader as an exercise.

Theorem 6.9. *If a ring has a multiplicative identity, then that multiplicative identity is unique.*

Now that we know that the additive and multiplicative identities of a ring are unique, we will usually label these identities as 0 and 1, respectively.

Our next theorem shows us that the property from the integers that multiplying any number by zero results is zero is not unique to our number systems, but is inherent in the properties of the additive identity in a ring.

Theorem 6.10. *If $\langle R, +, \cdot \rangle$ is a ring, and if 0 is the additive identity in R, $0 \cdot a = a \cdot 0 = 0$ for all $a \in R$.*

Proof. Let $\langle R, +, \cdot \rangle$ be a ring with additive identity 0 and let $a \in R$. Then $a \cdot 0 = a \cdot (0 + 0)$, since 0 is the additive identity. From the distribution property we have that $a \cdot 0 = a \cdot 0 + a \cdot 0$ and if we add the additive inverse of $a \cdot 0$ to both sides of the equation, we have that $0 = a \cdot 0$. We can similarly show that $0 \cdot a = 0$. $\square$

Recall from Theorem 6.2 that we have uniqueness of the additive inverse since $\langle R, +\rangle$ is an abelian group. We can also prove the uniqueness of a multiplicative inverse, if such an inverse exists because in this situation, $\langle R, \cdot\rangle$ forms a group.

One question that students often ask when they are introduced to multiplication with negative numbers why a negative times a negative is a positive. We see from the theorem below that this property, and other related properties are inherited from the ring structure of the number system. However, when explaining this property to students, it is often better to focus on the symmetry about the additive identity. In particular, with integers and real numbers it is good to focus on the symmetry of the additive identities on the number line. For the complex numbers it is helpful to focus the students' attention on the symmetry about the additive identity when looking at the complex plane.

Theorem 6.11. *If $\langle R, +, \cdot\rangle$ is a ring and $a, b \in R$, then*

- $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$
- $(-a) \cdot (-b) = a \cdot b$

Proof. In order to prove the first part of the theorem we need to only show that

$$(a \cdot (-b)) + (a \cdot b) = 0 \quad \text{and} \quad ((-a) \cdot b) + (a \cdot b) = 0$$

because the additive inverse of $a \cdot b$ is unique.

We see that

$$\begin{aligned}(a \cdot (-b)) + (a \cdot b) &= a \cdot ((-b) + b) \text{ (from the distributive property)}\\ &= a \cdot 0 \text{ (since } -b \text{ is the additive inverse of } b)\\ &= 0\end{aligned}$$

as a result of Theorem 6.10. Similarly,

$$\begin{aligned}((-a) \cdot b) + (a \cdot b) &= ((-a) + a) \cdot b \text{ (from the distributive property)}\\ &= 0 \cdot b \text{ (since } -a \text{ is the additive inverse of } a)\\ &= 0\end{aligned}$$

as a result of Theorem 6.10. For the second result we use the first result to see that

$$(-a) \cdot (-b) = -(a \cdot (-b)) = -(-(a \cdot b)) = a \cdot b$$

since the additive inverse of the additive inverse of an element is the original element. □

6.3.5 Integral Domains

We often want to solve an equation such as $ax=b $ or $xa=b $ for x with $a \neq 0$. In order to do so we need to use the cancellation laws, $ab = ac$ with $a \neq 0$ implies $b = c$ and $ba = ca$ with $a \neq 0$ implies $b = c$.

We will see that the cancellation laws correspond with a property about zero divisors.

Definition 6.7. If a and b are two non-zero elements of a ring R such that $ab = 0$, then a and b are **divisors of** 0 (or **zero divisors**). In particular, a is a **left divisor** of 0 and b is a **right divisor** of 0.

Theorem 6.12. *The cancellation laws hold in a ring R if and only if R has no left or right divisors of* 0.

Proof. Assume that for all elements a, b, and c of a ring R such that $a \neq 0$ we have that

$$(ab = ac) \Rightarrow (b = c).$$

We can then add the additive inverse of ac to both sides of the first equation and the additive inverse of c to both sides of the right equation to see that this implication is equivalent to

$$((ab - ac) = 0) \Rightarrow (b - c = 0).$$

Using the left distribution property of the ring, we have that the implication is equivalent to

$$(a \cdot (b - c) = 0) \Rightarrow ((b - c) = 0)\,.$$

This implication is true if and only if a is not a left divisor of 0.

Similarly, one can show that the right cancellation law of

$$(ba = ca, \;\text{ with } a \neq 0) \Rightarrow (b = c)$$

is equivalent to a not being a right divisor of 0.

Therefore, we see that the cancellation laws holding for a ring is equivalent to the ring having no zero divisors. ☐

Since this property of not having zero divisors is so important in the process of solving simple equations, we give rings with this property a label.

Definition 6.8. An **integral domain** is a commutative ring with unity containing no divisors of 0.

A very useful example of an integral domain is the set of integers with the usual addition and multiplication. We will see in the next chapter that the polynomials also form an integral domain, which allows us to factor polynomials in a similar way to factoring integers.

In order to better understand how all of these different types of rings fit together, Figure 6.1 is a graphical way to understand their relationships.

6.3.6 Ring Homomorphisms

Just like group homorphisms in Section 6.1.3, we can study the functions between rings that maintain the ring structures.

Definition 6.9. Let $\langle R, +, \cdot\rangle$ and $\langle R', *, \times\rangle$ be rings. A map $\phi : R \to R'$ is a **ring homomorphism** if

- $\phi(a + b) = \phi(a) * \phi(b)$ for all a and b in R
- $\phi(a \cdot b) = \phi(a) \times \phi(b)$ for all a and b in R.

ϕ is called a **ring isomorphism** if it is also a bijection. If such an isomorphism exists from a ring $\langle R, +, \cdot\rangle$ to a ring $\langle R', *, \times\rangle$, then we say that $\langle R, +, \cdot\rangle$ is **isomorphic** to $\langle R', *, \times\rangle$.

When we talk about rings, we often refer to a ring $\langle R, +, \cdot\rangle$ as just R when the binary operations are easily inferred from the set. For instance, we often talk about $\mathbb{R}$ as the ring of real numbers with the binary operations of addition and multiplication being inferred.

Similar to groups, it is useful to know when two rings are really the “same” ring with different labels. We see that just like with groups (Theorem 6.7) we see that ring isomorphisms create an equivalence relation and so two rings are the “same” if there is an isomorphism from one to the other. Therefore, in order to prove that two rings are isomorphic, it is usually done by finding the isomorphism. However, to prove that two rings are not isomorphic we usually use a proof by contradiction using certain properties of ring isomorphisms.

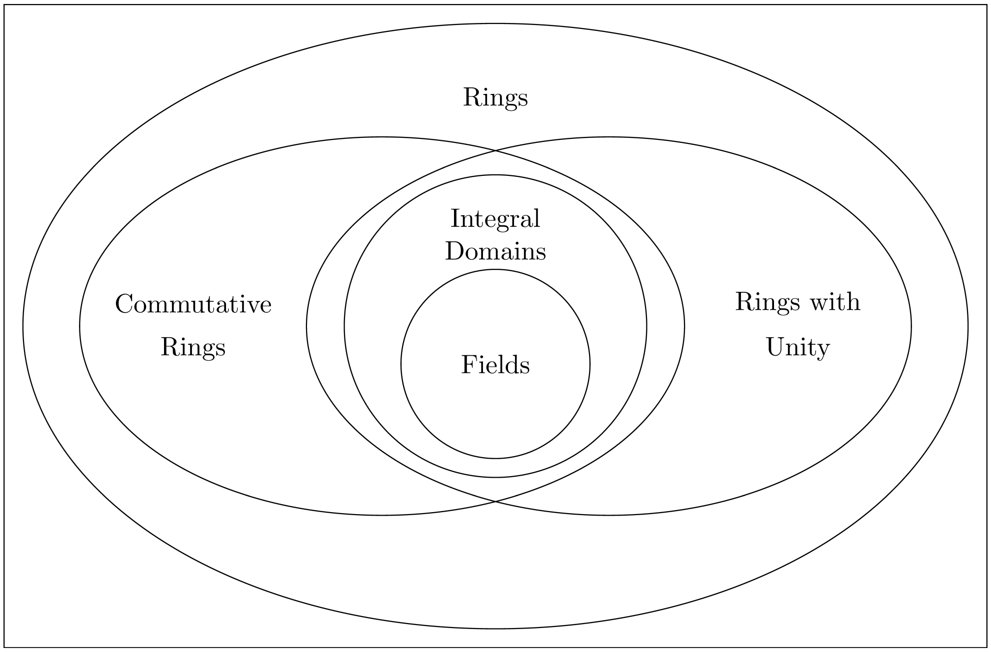


Figure 6.1: Venn diagram of ring relationships

6.3.7 Exercises

1. Create Cayley tables for addition and multiplication for $\mathbb{Z}_{12}$. What patterns do you recognize inside of the table and how would you explain the reasons for those patterns?

2. What are the zero divisors of $GL(2, \mathbb{Z}_2)$?

3. For each of the following rings, determine the appropriate location for the ring in the Venn diagram of the ring relationships (Figure 6.1).

   a. $\langle \mathbb{Z}, +, \cdot \rangle$, the integers with the usual addition and multiplication.
   b. $\langle \mathbb{Q}, +, \cdot \rangle$, the rational numbers with the usual addition and multiplication.
   c. $\langle \mathbb{R}, +, \cdot \rangle$, the real numbers with the usual addition and multiplication.
   d. $\langle \mathbb{C}, +, \cdot \rangle$, the complex numbers with the usual addition and multiplication.
   e. $\langle 2\mathbb{Z}, +, \cdot \rangle$, the even integers with the usual addition and multiplication.
   f. $\langle \mathbb{Z}_2, +, \cdot \rangle$, as defined in Example 6.8.
   g. $\langle \mathbb{Z}_4, +, \cdot \rangle$, as defined in Example 6.10.
   h. $\langle GL(2, \mathbb{R}), +, \cdot \rangle$, the two by two matrices with real coefficients with the usual matrix addition and multiplication.
   i. $\langle GL(2, 2\mathbb{Z}), +, \cdot \rangle$, the two by two matrices with even integer coefficients with the usual matrix addition and multiplication.

4. Prove Theorem 6.9.

5. Show that $a^2 - b^2 = (a + b)(a - b)$ for all a and b in R if and only if R is a commutative ring.

6. Prove that $2\mathbb{Z}$ and $3\mathbb{Z}$ are not isomorphic as rings.

7. Let R be a ring and let R^R be the set of all functions mapping R into R. For $f, g \in R^R$, define the sum $f + g$ by
$$(f + g)(a) = f(a) + g(a)$$

and the product $f \cdot g$ by

$$(f \cdot g)(a) = f(a) \cdot g(a)$$

for each $a \in R$. Show that $\langle R^R, +, \cdot \rangle$ is a ring.

8. Vector spaces, or specifically Euclidean spaces, constitute a subject of significant study in secondary and tertiary mathematics.

   a. Discuss how vector spaces are related to groups, rings, and fields by describing their similarities differences or inclusions.
   b. In what ways are linear transformations between vector spaces related to homomorphisms of groups or rings.

Chapter 7

Integral Domains and Polynomials

7.1 Properties of the Ring of the Integers

In this section we spend time better understanding the integral domain of the integers. We will look at some of the divisibility properties that are the underpinnings for our standard long division algorithms and build up to the prime factorization of the integers.

The first key component of the integers that we will look at is the Division Algorithm that we first stated and proved in Section 4.3.

Theorem 7.1 (The Division Algorithm for Integers)**.** *If a and b are integers with $b > 0$, then there are unique integers q and r such that*

$$a = bq + r, \quad \textit{with } 0 \leq r < b.$$

We call the integer q in the division algorithm the quotient of a and b and we call r the remainder of the quotient of a and b.

Related Content Standards

- (4.NBT.6) Find whole-number quotients and remainders with up to four-digit dividends and one-digit divisors, using strategies based on place value, the properties of operations, and/or the relationship between multiplication and division. Illustrate and explain the calculation by using equations, rectangular arrays, and/or area models.
- (5.NBT.6) Find whole-number quotients of whole numbers with up to four-digit dividends and two-digit divisors, using strategies based on place value, the properties of operations, and/or the relationship between multiplication and division. Illustrate and explain the calculation by using equations, rectangular arrays, and/or area models.

Proof. Let a and b be integers with $b > 0$. We will first prove the existence of integers q and r such that

$$a = bq + r, \quad \text{with } 0 \leq r < b$$

and then prove that these integers are unique.

- **Existence.** If $a \geq 0$, then we can let

$$S = \{a - kb |\, a - kb \geq 0 \text{ and } k \text{ is a non-negative integer}\}.$$

Since $a \geq 0$, we see that $a = a - 0 \cdot b$ is an element of S and so S is non-empty. We also see that $S \subseteq [0, a]$.

So by the Well-Ordering Property of the Integers, there is a smallest element of S which we will call r. Since $r \in S$, there exists a non-negative integer q such that $r = a - qb$.

Since r is the smallest element of S and since $a - (q + 1)b < a - qb = r$, we see that $a - (q + 1)b < 0$, otherwise it would be a smaller element of S.

This implies that $r = a - qb < b$. Therefore we have the existence of the q and r.

If $a < 0$, using the above process we can find non-negative integers q and r, with $0 \leq r < b$, such that $-a = bq + r$. So $a = b(-q) + r$ satisfies the existence statement.

- **Uniqueness.** In order to prove the uniqueness of the quotient and remainder, we assume that there are two pairs of integers (q_1, r_1) and (q_2, r_2) such that

$$a = bq_1 + r_1 \text{ and } a = bq_2 + r_2, \text{ with } 0 \leq r_1 < b \text{ and } 0 \leq r_2 < b.$$

This implies that $bq_1 + r_1 = bq_2 + r_2$ and by rearranging the terms we have that

$$b(q_1 - q_2) = r_2 - r_1.$$

If $0 \leq r_1 \leq r_2 < b$, then $0 \leq b(q_1 - q_2) = r_2 - r_1 < b$. So $q_1 - q_2$ is an integer in $[0, 1)$ and so must be 0. Therefore, $q_1 = q_2$ and so $r_1 = r_2$. Therefore, the quotient and remainder are unique.

□

It is through a repetition of a variation of this division algorithm that we are able to divide integers with the standard long division algorithm. Look at various ways that we can find the quotient of 284 by 13. In particular compare the division algorithm in Theorem 7.1, the standard long division algorithm, the area model, and other models with which you are familiar.

7.1.1 Composing and Decomposing Integers

Numerical fluency involves the ease at which one is able to break apart and recombined numbers under addition and/or multiplication. For addition, the fluency focuses on the tens using items like ten frames. For multiplication, it is a fluency with multiplication facts up to 12s, with flexibility and reversibility.

Related Content Standards

- (4.OA.4) Find all factor pairs for a whole number in the range 1-100. Recognize that a whole number is a multiple of each of its factors. Determine whether a given whole number in the range 1-100 is a multiple of a given one-digit number. Determine whether a given whole number in the range 1-100 is prime or composite.

Definition 7.1. Let a and b be integers with $b \neq 0$. We say that b **divides** a, or b is a **divisor** of a, denoted by $b|a$, if there exists a unique $q \in \mathbb{Z}$ such that $a = bq$. So b divides a if and only if a **is a multiple of** b.

We will look at the various properties of the divisors of integers later in this section. But first, we will look at those integers for which there are no divisors other than 1 and itself.

Definition 7.2. A natural number $p > 1$ is **prime** if the only positive divisors of p are 1 and p. A natural number $n > 1$ is **composite** if it is not prime.

Note that 1 is not a prime number, since prime numbers are defined to be integers greater than 1. The primary reason behind this is to make the prime factorization of the integers unique.

So the prime numbers less than 100 are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, and 97.

Prime numbers have been studied for centuries, with Euclid proving that there are an infinite number of prime numbers in Proposition 20 of Book IX of the *Elements*. Over the centuries, many different mathematicians have proven this using various techniques that include the study of Fermat numbers ($F_n = 2^{2^n} + 1$), group theory, and topology. However, many of the proofs are based on the following lemmas.

Lemma 7.1. *If n is a composite, then n is divisible by a prime number.*

Proof. Assume that n is a composite number that is not divisible by a prime number. So there are natural numbers n_1 and n_2, both larger than 1, such that $n = n_1 \cdot \tilde{n}_1$. From the properties of multiplication of the natural numbers we know that $n_1 < n$ and $\tilde{n}_1 < n$. Since n is not divisible by a prime number, n_1 is a composite number and so we can further factor n_1 as a product of natural numbers, n_2 and $\tilde{n}_2$. This process will create a sequence of composite numbers,

$$n > n_1 > n_2 > \dots$$

. The set of such n_i is a bounded subset of the natural numbers and by the well-ordering property (3.4), there is a least element, m. However, this number would be both composite and prime. Since this is not possible, we see that n is divisible by a prime number. □

We are now ready to prove the infinitude of the prime numbers.

Theorem 7.2. *There are infinitely many prime numbers.*

Since we define infinitely many as the negation of finitely many, we will use the technique of "proof by contradition."

Proof. Assume that there are finitely many prime numbers,

$$\{p_1, p_2, p_3, \cdots, p_n\}$$

with $p_i < p_{i+1}$ for each $1 \le i \le n$. So we can define a natural number

$$m = p_1 \cdot p_2 \cdots p_n + 1 = \left(\prod_{i=1}^{n} p_i \right) + 1.$$

So for each $1 \le i \le n$ we have that

$$m = p_i \cdot \left(\prod_{j \ne i} p_j \right) + 1$$

and so by the division algorithm we see that m is not divisible by any of the prime numbers, p_i. However, our previous lemma states that this is not possible. Therefore, there are infinitely many prime numbers. □

7.1.2 Greatest Common Divisor and Least Common Multiple

Definition 7.3. Let a and b be integers, not both zero. We say that the **greatest common divisor** (also called the **greatest common factor**) of a and b, denoted $\gcd(a, b)$, is the natural number d such that

- d divides both a and b, and
- if u divides both a and b for some integer u, then $u \le d$.

Let a and b be integers, not both zero. We say that the **least common multiple** of a and b, denoted $\text{lcm}(a, b)$, is the natural number m such that

- m is a multiple of both a and b, and
- if u is a multiple of both a and b for some integer u, then $u \geq m$.

Related Content Standards

- (6.NS.4) Find the greatest common factor of two whole numbers less than or equal to 100 and the least common multiple of two whole numbers less than or equal to 12. Use the distributive property to express a sum of two whole numbers 1-100 with a common factor as a multiple of a sum of two whole numbers with no common factor.

There are many different options for finding the greatest common divisor and least common multiple of two numbers. If the numbers are small, one can find the greatest common divisor by listing out all of the possible factors of each number and choosing the largest one in common. Another option is using the factor trees of the two numbers, based on the prime factorizations of the numbers. Similarly, for small numbers one can list out the multiples of each number until you find a common multiple, or you can use information from the numbers' prime factorizations. When the numbers are much larger, one can develop other techniques to create algorithms for finding the greatest common divisor and least common multiple. Many of these techniques are based on the following theorem.

Theorem 7.3. *If a, b, q, r are integers, with a and b not both zero, and if*

$$a = bq + r,$$

then $\gcd(a, b) = \gcd(b, r)$.

Proof. Since a and b are not both zero, we let $d = \gcd(a, b)$. This means that there are integers m and n such that $a = md$ and $b = nd$. Therefore,

$$r = a - bq = md - (nd)q = d \cdot (m - nq)$$

and so d divides r and d divides b. Since d is a common divisor of b and r, we see that

$$\gcd(a, b) \leq \gcd(b, r).$$

Since $a = bq + r$, any common divisor of b and r must also be a divisor of a, we have that $\gcd(b, r)$ is a common divisor of a and b and so

$$\gcd(b, r) \leq \gcd(a, b).$$

Therefore, $\gcd(a, b) = \gcd(b, r)$. □

Theorem 7.4 (Euclidean Algorithm)**.** *Let a and b be two positive integers with $a > b$. Then the greatest common divisor of a and b can be found using the following algorithm.*

- *[Step 1] Apply the division algorithm to find integers q and r such that $a = bq + r$ with $0 \leq r < b$.*
- *[Step 2] If $r = 0$, then b is $\gcd(a, b)$.*
- *[Step 3] While $r > 0$, replace define $a := b$ and $b := r$ and apply the division algorithm to the new a and b to generate a new pair of integers q and r. Once $r = 0$, then the value of b is the greatest common divisor of the original integers a and b.*

In order to better understand the algorithm, we will find the greatest common divisor of 924 and 260. We start by letting $a = 924$ and $b = 260$. From the division algorithm we find that

$$924 = 3 \cdot 260 + 144.$$

We now let $a = 260$ and $b = 144$ and use the division algorithm again to find that

$$260 = 1 \cdot 144 + 116.$$

Continuing this process we see that $\gcd(924, 260) = 4$.

$$\begin{aligned} 144 &= 1 \cdot 116 + 28 \\ 116 &= 4 \cdot 28 + 4 \\ 28 &= 7 \cdot 4 + 0 \end{aligned}$$

7.1.3 Fundamental Theorem of Arithmetic

The most common way to find the greatest common divisor and least common multiple of two numbers is dependent upon the numbers' prime factorizations. This ability to write a number uniquely as a product of primes is called the Fundamental Theorem of Arithmetic, which we will prove in this section.

We say that two numbers are **relatively prime** if their greatest common divisor is 1.

Theorem 7.5. *Natural numbers a and b are relatively prime if and only if there exist integers s and t such that $sa + tb = 1$.*

Proof. Suppose that a and b are relatively prime. From the proof of the Euclidean algorithm we can see that for integers a and b that there exists integers s and t such that $sa + tb = \gcd(a, b) = 1$.

Conversely, if there exist integers s and t such that $sa + tb = 1$, we know that $d = \gcd(a, b)$ divides a and b and so would also divide 1. Since the only integers that divide 1 are 1 and -1 and since $\gcd(a, b) \geq 1$, we have that $\gcd(a, b) = 1$. □

Lemma 7.2. *If p is a prime number and a and b are integers such that p divides ab, then p divides a or p divides b.*

Proof. In order to prove the Lemma, we will prove that the negation is false. Suppose that a is relatively prime to p and b is relatively prime to p. Then from Theorem 7.5 we have integers s, t, u, and v such that $sa + tp = 1$ and $ub + vp = 1$. Then

$$1 = 1 \cdot 1 = (sa + tp) \cdot (ub + vp) = (su)(ab) + (sav + tub + tvp)p$$

and so ab and p are relatively prime, implying that p does not divide ab. □

Theorem 7.6 (Fundamental Theorem of Arithmetic)**.** *Every integer $n > 1$ can be factored as a product of prime numbers, and except for the order of the factors, this prime factorization is unique.*

Proof. Let $n > 1$ be an integer. If n is prime, we have a prime factorization of n. If n is not prime, then there are integers n_1 and n_2 such that $n = n_1 n_2$ and we can assume without loss of generality that $1 < n_1 \leq n_2 < n$. If either of these integers are composite, then they also can be factored into factors strictly smaller than them. This process can be continued, but it cannot be continued indefinitely since n is a finite number. At the conclusion of the process, n will have a prime factorization. We now only need to show uniqueness of this factorization.

Assume that n has two distinct prime factorizations,

$$n = p_1 p_2 \cdots p_j \quad \text{and} \quad n = q_1 q_2 \cdots q_k$$

and we can assume without loss of generality that $j \leq k$.

Since p_1 divides n, a generalization of Lemma 7.2 gives us that p_1 must divide one of the q_i. By reordering the primes we can label this q_i as q_1. Since both p_1 and q_1 are prime, they must be equal. We can continue this process with $p_2 \cdots p_j$ and $q_2 \cdots q_k$ to find that $p_2 = q_2$ and continuing the process to equate each of the p_i with a q_i. If some of the q_i remain after this process, these would be factors of 1 and therefore could not be prime. Therefore, $j = k$ and we see that the two prime factorizations are just a reordering of each other. ☐

7.1.4 Exercises

1. Let $n \in \mathbb{Z}$. What are the possible values for the remainder of n^2 with a divisor of 3? Justify your result.

2. Find the q and r from the Division Algorithm for the given values of a and b.

   a. $a = 0$, $b = 14$
   b. $a = -42$, $b = 14$
   c. $a = 9$, $b = 14$

3. Use the Division Algorithm to prove that for positive integers m and n, with $m > n$, then exactly one of the following statements must be true:

   - There is a positive integer k such that $m = kn$
   - There is a positive integer k such that $kn < m < (k+1)n$

4. Use the Euclidean Algorithm to find $\gcd(4368, 31050)$.

5. Create a computer program or spreadsheet based on the Euclidean Algorithm that you can use to find the greatest common divisor of two integers.

6. Prove that for all integers that $\gcd(a, b) \cdot \operatorname{lcm}(a, b) = |ab|$. (Hint: Use the prime factorizations of a and b.)

7.2 Polynomial Rings

Another set of examples of integral domains in the secondary curriculum are polynomial rings. We let

$$\mathbb{R}[x] = \left\{ a(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = \sum_{j=0}^{n} a_j x^j \,\middle|\, a_j \in \mathbb{R} \right\}$$

The concept of a polynomial can be generalized to not just having real number coefficients. For instance, we will sometimes require that the polynomials have integer coefficients or possibly coefficients that are complex numbers. So we will generalize the definition of a polynomial to be able to be used in these cases.

Definition 7.4. Let R be a ring. We define a polynomial $f(x)$ with coefficients in R to be an infinite formal sum

$$\sum_{j=0}^{\infty} a_j x^j = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n + \cdots$$

where $a_j \in R$ for all j and $a_j = 0$ for all but finitely many values of j.

We denote the set of such polynomials as $R[x]$.

Since the $a_j = 0$ for all but finitely many values of j, by the well-ordering principle, there is a largest value of j for which $a_j \neq 0$. We call this largest value of j the **degree** of the polynomial. If all of the coefficients are 0, we say the polynomial is degree zero.

So we can define the function "degree" that maps the non-zero polynomials to the natural numbers by

$$\text{degree}\left(\sum_{j=0}^{\infty} a_j x^j\right) \text{ is the largest value of } j \text{ for which } a_j \text{ is non-zero.}$$

We call the a_j that corresponds to this value of j the **leading coefficient**.

Related Content Standards

- (HSA.APR.1) Understand that polynomials form a system analogous to the integers, namely, they are closed under the operations of addition, subtraction, and multiplication; add, subtract, and multiply polynomials.

If $a(x) = \sum_{j=0}^{n} a_j x^j$ and $b(x) = \sum_{j=0}^{m} b_j x^j$ are two elements of $R[x]$, with $n \leq m$, then we define

$$a(x) + b(x) := \sum_{j=0}^{m} (a_j + b_j) x^j$$

where $a_j := 0$ for $n < j \leq m$. From this definition, we see that addition is closed. We also see that addition is associative from the associativity of addition on R. Since R is a ring, addition is commutative on R and so addition is commutative on $R[x]$. We also see that the zero polynomial, $0(x) = 0$, is an additive identity for the set. So $(R[x], +)$ is an Abelian group.

Multiplication on $R[x]$ is more challenging to write in a compact form. We will assume that the indeterminate, x, commutes multiplicatively with every element in R, i.e. $ax = xa$, $\forall a \in R$ (the set of all elements of a ring that commute with every element of the ring is called the center of the ring). We can then define multiplication as

$$\begin{aligned} a(x) \cdot b(x) &:= \left(\sum_{j=0}^{n} a_j x^j\right) \cdot \left(\sum_{k=0}^{m} b_k x^k\right) = \sum_{j=0}^{n} \left(a_j x^j \left(\sum_{k=0}^{m} b_k x^k\right)\right) \\ &= \sum_{j=0}^{n} \sum_{k=0}^{m} \left(a_j x^j b_k x^k\right) = \sum_{j=0}^{n} \sum_{k=0}^{m} (a_j b_k)\, x^{(j+k)} \end{aligned}$$

and we see that $R[x]$ is closed under multiplication.

Unless R is a commutative ring, we must maintain the order of the coefficients. The associativity of multiplication and distributive properties are inherited from $R[x]$. If the ring R has a multiplicative identity, then $R[x]$ also has a multiplicative identity of the polynomial of the multiplicative identity. Finally, if R is a commutative ring, then multiplication on $R[x]$ is commutative.

Theorem 7.7. *The set $R[x]$ of polynomials with coefficients from a ring R with the binary operations of polynomial addition and multiplication is a ring.*

One of the main activities that we do with polynomials in the secondary curriculum is to treat the polynomials as functions and evaluate them for various values.

The following theorem gives us the mathematical justification for such evaluations.

In this theorem we use a field F and a subfield E, with is a subset of F that also satisfies the field properties inherited from F. The reason that we use the field E and the subfield F is that we often have polynomials with whose coefficients are from one field and the point at which we are evaluating from a larger field. For instance, we often look at quadratic polynomials with real coefficients and determine that they have complex zeros. In this instance, the subfield would be $\mathbb{R}$ and the larger field would be $\mathbb{C}$.

Theorem 7.8. *Let F be a subfield of a field E and let α be any element of E. The map $\phi_\alpha : F[x] \to E$ defined by*

$$\phi_\alpha(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_0 + a_1\alpha + a_2\alpha^2 + \cdots + a_n\alpha^n$$

is a ring homomorphism from $F[x]$ into E.

We go ahead and include the proof of this theorem below to demonstrate how these evaluation homomorphisms work and to demonstrate certain properties of polynomial functions.

Proof. Let F be a subfield of a field E and let α be any element of E. We also let $a(x)$ and $b(x)$ be generic elements of $F[x]$ and label them as

$$a(x) = \sum_{j=0}^{\infty} a_j x^j \text{ and } b(x) = \sum_{j=0}^{\infty} b_j x^j$$

where the a_j and the b_j are zero for all but finitely many values of j.

Then $\phi_\alpha(a(x)) = \sum_{j=0}^{\infty} a_j\alpha^j$ and $\phi_\alpha((b(x))) = \sum_{j=0}^{\infty} b_j\alpha^j$, which are elements of E since addition and multiplication on E are closed and only finitely many elements of each of the sums are non-zero.

Using the definition of addition of two polynomials, we have that $a(x) + b(x)$ is a polynomial that is written as

$$a(x) + b(x) = \sum_{j=0}^{\infty} (a_j + b_j)x^j.$$

Therefore,

$$\phi_\alpha(a(x) + b(x)) = \sum_{j=0}^{\infty} (a_j + b_j)\alpha^j$$

and by rearranging the finite number of terms using the distributive property and that both addition and multiplication are commutative in E we have that

$$\phi_\alpha(a(x) + b(x)) = \sum_{j=0}^{\infty} a_j\alpha^j + \sum_{j=0}^{\infty} b_j\alpha^j = \phi_\alpha(a(x)) + \phi_\alpha(b(x)).$$

Similarly, one finds that

$$\phi_\alpha(a(x) \cdot b(x)) = \phi_\alpha(a(x)) \cdot \phi_\alpha(b(x)).$$

□

You should keep in mind that the operations are using the same symbol for operations in two different rings. For the case of $a(x) + b(x)$, the addition is in the polynomial ring $F[x]$, while $\phi_\alpha(a(x)) + \phi_\alpha(b(x))$ is addition in E.

We say that an element $a \in E$ is a **zero of a polynomial** $p(x) \in F[x]$ if

$$\phi_a(p(x)) = 0.$$

We would really like to have the property that for all $a, b \in R$, a and b are the only zeros of the polynomial $(x - a)(x - b)$. In order for that to be true, we need $R[x]$ to be an integral domain. However, we first need some properties related to the degree of polynomials.

Theorem 7.9. *Let R be an integral domain and let $a(x)$ and $b(x)$ be elements of $R[x]$. Then*

$$\text{degree}(a(x)+b(x)) \leq \max\left(\text{degree}(a(x)), \text{degree}(b(x))\right), \textit{ and}$$

$$\text{degree}\left(a(x)b(x)\right) = \text{degree}(a(x)) + \text{degree}(b(x)).$$

Proof. Let $a(x) = \sum_{j=0}^{n} a_j x^j \in R[x]$ with degree n and $b(x) = \sum_{j=0}^{m} b_j x^j \in R[x]$ with degree m. We can assume without any loss of generality that $n \leq m$.

If $n < m$, $a(x)+b(x) = \sum_{j=0}^{m}(a_j+b_j)x^j$, where $a_j = 0$ for $n < j \leq m$, we see that degree$(a(x)+b(x)) = m$. If $n = m$ the sum also has degree m, unless the two leading coefficients are additive inverses. Therefore,

$$\text{degree}(a(x)+b(x)) \leq \max\left(\text{degree}(a(x)), \text{degree}(b(x))\right).$$

Since

$$a(x)\cdot b(x) = \sum_{j=0}^{n}\sum_{k=0}^{m}\left(a_j b_k\right)x^{(j+k)}$$

we see that the leading term of $a(x)\cdot b(x)$ is $a_n b_m$. This is non-zero because R is an integral domain. So the degree of $a(x)\cdot b(x)$ is $m+n$. And we have that

$$\text{degree}\left(a(x)b(x)\right) = \text{degree}(a(x)) + \text{degree}(b(x)).$$

□

We are now ready to prove that R being an integral domain implies the $R[x]$ is an integral domain.

Theorem 7.10. *If R is an integral domain, then $R[x]$ is an integral domain.*

Proof. Assume that there are two polynomials, $a(x) = \sum_{j=0}^{n} a_j x^j$ and $b(x) = \sum_{j=0}^{m} b_j x^j$ in $R[x]$, such that $a_n \neq 0$ and $b_m \neq 0$. Then the product of the polynomials has degree of $n+m$. If $n+m > 0$, then the product cannot be the zero polynomial. If $n+m = 0$, then n and m are both zero and $a(x) = a_0$, $b(x) = b_0$, and $a(x)\cdot b(x) = a_0 b_0$. Since $a_0 \neq 0$ and $b_0 \neq 0$, since R is an integral domain, the product is non-zero. □

7.2.1 Exercises

1. In $\mathbb{Z}_{12}$, what are the zeros of the polynomial $6 - 5x + x^2$?

2. Use the various models of multiplication presented in Section 4.2 to add and multiply the following pairs of polynomials:

   a. $a(x) = -2x + 3$ and $b(x) = x - 1$
   b. $a(x) = 2x^2 - 4x + 1$ and $b(x) = x^3 - 2$
   c. Which models transition well to polynomials and which do not?

7.3 Properties of Polynomial Rings

The properties of irreducible (prime) elements and factorizations is very important in the study of the integers and is dependent upon the integral domain properties. We have also seen that certain polynomial rings are also integral domains. In this section we will see how the factorization of integers into prime factors can be extended to the ring of polynomials. Throughout this section we will let S be a general integral domain (you can think of it as $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$) and let $S[x]$ be the ring of polynomials with coefficients from S.

7.3.1 Polynomial Division

We begin this generalization by finding a generalization of the division algorithm. We see from the Division Algorithm for Integers that we need to have some type of ordering for the elements of our integral domain. We will call such an ordering a Euclidean function.

Definition 7.5. Let R be an integral domain. A **Euclidean function** on R is a function f from $R \setminus \{0\}$ to the non-negative integers satisfying the fundamental division-with-remainder property.

- If a and b are in R and b is non-zero, then there exist q and r in R such that $a = bq + r$ and either $r = 0$ or $f(r) < f(b)$.

An integral domain for which a Euclidean function exists is called a Euclidean domain.

Since S is an integral domain, we see that $S[x]$ is also an integral domain (Theorem 7.10). We can now use the degree function of the polynomial as a Euclidean function and state a Division Algorithm for Polynomials that makes $S[x]$ a Euclidean domain, when S is a field.

Theorem 7.11 (Division Algorithm for Polynomials)**.** *Let S be a field. Suppose that $a(x)$ and $b(x)$ are polynomials in $S[x]$ and that $b(x)$ is not the zero polynomial. Then there are unique $q(x)$ and $r(x)$ in $S[x]$ such that*

$$a(x) = b(x) \cdot q(x) + r(x)$$

and $r(x)$ is the zero polynomial or the degree of $r(x)$ is less than the degree of $b(x)$.

The reason that we need S to be a field is to allow for division of the coefficients in the division algorithm. For example in $\mathbb{Z}[x]$, if $a(x) = x^2$ and $b(x) = 3x$, the $r(x)$ in the division algorithm would need to be $3x$.

Proof. Let S be a field and let $a(x)$ and $b(x)$ be polynomials, with $b(x)$ not the zero polynomial.

Let

$$T = \{a(x) - b(x) \cdot s(x) | s(x) \in S[x]\}$$

and from the well-ordering property we can choose an $r(x)$ in T with minimal degree. So

$$a(x) = b(x)q(x) + r(x)$$

for some $q(x) \in S[x]$.

Let $m = \text{degree}(b(x))$ and $n = \text{degree}(r(x))$, and let b_m be the leading coefficient of $b(x)$ and r_n be the leading coefficient of $r(x)$. If $m \geq n$,

$$a(x) - b(x)\left(q(x) - \frac{r_n}{b_m}x^{m-n}\right) = r(x) - \frac{r_n}{b_m}x^{m-n}b(x)$$

which is an element of T with lower degree than $r(x)$, contradicting the choice of $r(x)$ to be an element of T with minimal degree. Therefore, $\text{degree}(r(x)) < \text{degree}(b(x))$.

For the uniqueness of $q(x)$ and $r(x)$, we assume that there are two pairs of such functions such that

$$a(x) = b(x)q_1(x) + r_1(x) \quad \text{and} \quad a(x) = b(x)q_2(x) + r_2(x)$$

with the degrees of $r_1(x)$ and $r_2(x)$ being less than the degree of $b(x)$. Subtracting the two equations and rearranging terms gives us

$$b(x)\left(q_1(x) - q_2(x)\right) = r_2(x) - r_1(x).$$

Since the degree of the polynomial on the right side of the equation is less than the degree of $b(x)$, $(q_1(x) - q_2(x))$ must be the zero polynomial. So $q_1(x) = q_2(x)$ and thus implying that $r_1(x) = r_2(x)$. And we have proven the uniqueness of the $q(x)$ and $r(x)$. $\square$

Related Content Standards

- (HSA.APR.6) Rewrite simple rational expressions in different forms; write $\frac{a(x)}{b(x)}$ in the form $q(x)+\frac{r(x)}{b(x)}$, where $a(x)$, $b(x)$, $q(x)$, and $r(x)$ are polynomials with the degree of $r(x)$ less than the degree of $b(x)$, using inspection, long division, or, for the more complicated examples, a computer algebra system.

Take some time and work through some examples of long division of polynomials and compare these with the procedures in the long division of integers. It is through these connections between the two processes that students are able to understand and retain the knowledge of the long division algorithm for polynomials.

We can now extend the properties of factors and prime (or irreducible) elements of the integers to the polynomial ring.

Definition 7.6. Let S be an ring and $S[x]$ be the corresponding ring of polynomials. Let $a(x)$ and $b(x)$ be polynomials in $S[x]$ with $b(x)$ not the zero polynomial. We say that $b(x)$ divides $a(x)$ if there exists a $q(x)$ in $S[x]$ such that $a(x) = b(x) \cdot q(x)$. In this situation we also say that

- $a(x)$ is a multiple of $b(x)$,
- $b(x)$ is a divisor of $a(x)$, and
- $b(x)$ is a factor of $a(x)$.

We can now state the prove the Remainder Theorem and Factor Theorem for polynomials.

Theorem 7.12 (Remainder Theorem)**.** *Let S be a field ($\mathbb{Q}, \mathbb{R}, \mathbb{C}$) and suppose that $a(x)$ is a polynomial of degree greater than or equal to 1 in $S[x]$, and let $c \in S$. Then $a(c) := \phi_c(a(x))$ is the remainder obtained by dividing $a(x)$ by $x-c$, where ϕ_c is the evaluation function.*

Proof. Applying the Division Algorithm to $a(x)$ and $x-c$ we see that there is a unique $q(x)$ and $r(x)$ such that $a(x) = q(x) \cdot (x-c) + r(x)$ with the degree of $r(x)$ less than the degree of $x-c$. Therefore, $r(x)$ is a constant polynomial that we will call r. Since the evaluation function is a ring homomorphism we have that

$$a(c) := \phi_c(a(x)) = \phi_c(q(x)) \cdot \phi_c(x-c) + \phi_c(r).$$

Since $\phi_c(x-c) = c-c = 0$ and since $0 \cdot \phi_c(q(x)) = 0$, we have that $a(c)$ is the remainder. $\square$

This leads us directly to the Factor Theorem since $a(x) = q(x) \cdot (x-c) + a(c)$.

Theorem 7.13 (Factor Theorem)**.** *Let S be a field and suppose that $a(x)$ is a polynomial of degree at least 1. Then $x-c$ is a factor of $a(x)$ if and only if $a(c) = 0$.*

Related Content Standards

- (HSA.APR.2) Know and apply the Remainder Theorem: For a polynomial $p(x)$ and a number a, the remainder on division by $x-a$ is $p(a)$, so $p(a) = 0$ if and only if $(x-a)$ is a factor of $p(x)$.

7.3.2 GCF and LCM for Polynomials

Definition 7.7. Let $a(x)$ and $b(x)$ be in $S[x]$, not both the zero polynomial. We say that a **greatest common divisor** (also called the **greatest common factor**) of $a(x)$ and $b(x)$ is a polynomial $d(x)$ such that

- $d(x)$ divides both $a(x)$ and $b(x)$, and

- if $u(x)$ divides both $a(x)$ and $b(x)$ for some polynomial $u(x)$, then $u(x)$ divides $d(x)$.

Note that there is not a single greatest common divisor of two polynomials since constant multiples of a greatest common divisor may also be greatest common divisors. In the situation that S is a field, we can multiply a greatest common divisor by the multiplicative inverse of its leading coefficient and we now have a monic polynomial (a polynomial with leading coefficient of the multiplicative identity) that is a greatest common divisor. This monic polynomial is unique and so we define $\gcd(a(x), b(x))$ to be the **unique monic greatest common divisor of** $a(x)$ **and** $b(x)$.

Following the line of proofs of the integers, we can prove a theorem analogous to Theorem 7.3 for polynomials.

Theorem 7.14. *If $a(x)$, $b(x)$, $q(x)$, and $r(x)$ are polynomials in $S[x]$ for a field S, with $a(x)$ and $b(x)$ not both the zero polynomial, and if*

$$a(x) = b(x)q(x) + r(x),$$

then $\gcd(a(x), b(x)) = \gcd(b(x), r(x))$.

This theorem is the basis for a generalization of the Euclidean Algorithm to the polynomials.

Theorem 7.15 (Euclidean Algorithm for Polynomials)**.** *Let $a(x)$ and $b(x)$ be two polynomials in $S[x]$ for a field S, with* $\text{degree}(a(x)) > \text{degree}(b(x))$. *Then the greatest common divisor of $a(x)$ and $b(x)$ can be found using the following algorithm.*

- *[Step 1] Apply the division algorithm to find polynomials $q(x)$ and $r(x)$ such that $a(x) = b(x)q(x)+r(x)$ with* $0 \leq \text{degree}(r(x)) < \text{degree}(b(x))$.
- *[Step 2] If $r(x)$ is the zero polynomial, then is* $\gcd(a, b)$ *is $b(x)$ times the multiplicative inverse in S of its leading coefficient.*
- *[Step 3] While $r(x)$ is not the zero polynomial, replace $a(x) := b(x)$ and $b(x) := r(x)$ and apply the division algorithm to the new $a(x)$ and $b(x)$ to generate a new pair of polynomials $q(x)$ and $r(x)$. Once $r(x)$ is the zero polynomial, the value of $b(x)$ is a greatest common divisor of the original polynomials $a(x)$ and $b(x)$ and* $\gcd(a, b)$ *is this polynomial times the multiplicative inverse in S of its leading coefficient.*

In order to better understand the Euclidean Algorithm for Polynomials we will use the algorithm to find the unique monic greatest common divisor of $a(x) = 2x^5 - x^4 - 41x^3 + 59x^2 - 43x + 60$ and $b(x) = x^3 + 2x^2 + x + 2$.

The first step is to use the division algorithm to find the quotient and remainder of $a(x)$ divided by $b(x)$. One of the most efficient methods of finding this quotient and remainder (other than using a computer algebra system) is through long division of polynomials.

$$
\begin{array}{r|rrrrrr}
 & & & & 2x^2 & -5x & -33\\ \hline
x^3+2x^2+x+2 & 2x^5 & -x^4 & -41x^3 & +59x^2 & -43x & +60\\
 & -2x^5 & -4x^4 & -2x^3 & -4x^2 & & \\ \hline
 & & -5x^4 & -43x^3 & +55x^2 & -43x & \\
 & & 5x^4 & +10x^3 & +5x^2 & +10x & \\ \hline
 & & & -33x^3 & +60x^2 & -33x & +60\\
 & & & 33x^3 & +66x^2 & +33x & +66\\ \hline
 & & & & 126x^2 & & +126
\end{array}
$$

Therefore,

$$a(x) = b(x) \cdot (2x^2 - 5x - 33) + (126x^2 + 126)$$

and so we know that $\gcd(a(x), b(x)) = \gcd(x^3 + 2x^2 + x + 2, 126x^2 + 126)$. Using long division of polynomials we can find the new quotient and remainder for these two polynomials.

$$
\begin{array}{r}
\frac{1}{126}x + \frac{1}{63} \\
126x^2+126 \overline{\big) \; x^3 + 2x^2 + x \;\; + 2} \\
\underline{-x^3 \qquad\quad -x \qquad} \\
2x^2 \qquad + 2 \\
\underline{-2x^2 \qquad - 2} \\
0
\end{array}
$$

So,

$$x^3 + 2x^2 + x + 2 = (126x^2 + 126)\left(\frac{1}{126}x + \frac{1}{63}\right) + 0$$

and we see that $126x^2 + 126$ is a greatest common divisor of $a(x)$ and $b(x)$. Therefore, the unique monic greatest common divisor of $a(x)$ and $b(x)$ is $x^2 + 1$.

7.3.3 Irreducible Polynomials

We now turn our attention to studying the analogy to prime numbers in the polynomial ring.

Definition 7.8. An element in an integral domain, D, is called a **unit** of D if it has a multiplicative inverse. A non-zero element, p, of D is called an *irreducible* of D if any factorization $p = ab$ in D either a or b is a unit.

When S is a field, we see that the units of $S[x]$ are the constant polynomials since the degree of the product of two polynomials in $S[x]$ is equal to the sum of the degrees of the factors.

Notice that since $\sqrt{2}$ is irrational that $x^2 - 2$ is irreducible in $\mathbb{Q}[x]$ but has the factorization of $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$ in $\mathbb{R}[x]$ and $\mathbb{C}[x]$. This means that the set of irreducible polynomials depends upon the field from which the coefficients reside.

The following is an essential theorem in understanding factorization of polynomials and is one of the main consequences of the concept of irreducible elements. However, the proof of the theorem requires building up significant machinery in abstract algebra and so we will not include it here.

Theorem 7.16. *Let $p(x)$ ba an irreducible polynomial in $S[x]$. If $p(x)$ divides $a(x)b(x)$ for $a(x), b(x) \in S[x]$, then either $p(x)$ divides $a(x)$ or $p(x)$ divides $b(x)$.*

With this property involving the factorization of polynomials that we use on a regular basis we can also prove a property similar to that of the Fundamental Theorem of Arithmetic for integers.

Theorem 7.17. *Let S be a field. Then every polynomial of positive degree in $S[x]$ can be represented as a product*

$$cp_1(x)p_2(x)\cdots p_k(x)$$

where $c \in S$ and the $p_i(x)$ are monic irreducible polynomials in $S[x]$. This representation is unique up to reordering the irreducible polynomials.

The proof of this theorem is very similar to the method used to prove the Fundamental Theorem of Arithmetic, and so we will not include the details here. However, this theorem demonstrates the importance for understanding the irreducible polynomials in $\mathbb{R}[x]$ and $\mathbb{C}[x]$ in the process of understanding the properties of polynomials with real and complex coefficients.

The main result in this area is the Fundamental Theorem of Algebra.

Theorem 7.18 (The Fundamental Theorem of Algebra)**.** *The irreducible monic polynomials in $\mathbb{C}[x]$ are of the form $x - c$.*

There are many different methods to prove this result, but the primary method uses the machinery of complex calculus.

Related Content Standards

- (HSN.CN.9) Know the Fundamental Theorem of Algebra; show that it is true for quadratic polynomials.

7.3.4 Exercises

1. Find a prime factorization of the following polynomials in $\mathbb{Z}[x]$, $\mathbb{Q}[x]$, $\mathbb{R}[x]$, and $\mathbb{C}[x]$.

   a. $4x^3 - 2x^2 + 2x - 1$
   b. $x^3 - 1$
   c. $6x^2 - 13x - 5$
   d. $x^5 - 2x^4 - 2x^3 + 4x^2 - 3x + 6$

2. Explain how the domain of the variable x affects the number of solutions to the equation

$$(x^4 - 1)(4x^4 - 9x^2 + 2) = 0.$$

3. Find the unique **monic** prime factorization of the polynomials in $\mathbb{Q}[x]$, $\mathbb{R}[x]$, and $\mathbb{C}[x]$.

   a. $x^4 + 4x^2 + 4$
   b. $7x^4 - 28$
   c. $2x^4 - 5x^2 - 3$
   d. $9x^4 + 440x^2 - 49$
   e. $x^{106} - x^{100}$
   f. $4x^4 + 12x^3 - 86x^2 - 92x - 48$
   g. $x^4 - x^2 - 2$
   h. $x^2 - \pi$
   i. $2x^4 - x^3 - 19x^2 + 2x + 30$
   j. $ax^2 + bx + c$ based on the various possible values of a, b, and c with $a, b, c \in \mathbb{Z}$.

4. Prove Theorem 7.14 using the proof of Theorem 7.3 as an outline.

5. Prove that if f is a Euclidean function on an integral domain R, then $f(a) \leq f(b)$ for all non-zero a and b in R.

6. Find the value c so that $(x - 3)$ is a factor of the polynomial $p(x)$.

$$p(x) = cx^3 - 15x - 68$$

7. Find the monic greatest common divisor of the polynomials $a(x) = 4x^5 + 12x^4 + 20x^3 + 16x^2 + 16x + 4$ and $b(x) = 6x^5 + 12x^4 + 18x^3 + 12x^2 + 12x$.

7.4 Rational Expressions

In the same way that one can extend the integers to the field of rational numbers by defining addition and multiplication on pairs of integers, one can extend the polynomial ring by defining addition and multiplication of pairs of polynomials and then defining a set of equivalence classes on this set.

Related Content Standards

- (HSA.APR.7) Understand that rational expressions form a system analogous to the rational numbers, closed under addition, subtraction, multiplication, and division by a non-zero rational expression; add, subtract, multiply, and divide rational expressions.

Let F be a field and let $F[x]$ be the integral domain of polynomials with coefficients from F. We define an equivalence relation on the set of rational expressions of polynomials as $\frac{p(x)}{q(x)}$ is equivalent to $\frac{a(x)}{b(x)}$ if and only if $p(x) \cdot b(x) = a(x) \cdot q(x)$.

Define $F(x)$ to be the set of equivalence classes of quotients of polynomials $\left[\frac{p(x)}{q(x)}\right]$ with $q(x)$ not being the zero polynomial under this equivalence relation. We then proceed with dropping the equivalence class notation to simplify the notation. We define the following operations on $F(x)$.

$$\frac{p(x)}{q(x)} + \frac{f(x)}{g(x)} = \frac{p(x) \cdot g(x) + q(x) \cdot f(x)}{q(x) \cdot g(x)} \quad \text{and} \quad \frac{p(x)}{q(x)} \cdot \frac{f(x)}{g(x)} = \frac{p(x) \cdot f(x)}{q(x) \cdot g(x)}$$

We call the set $F(x)$ the set of rational expressions with coefficients from F and we label the corresponding set with operations as $\langle F(x), +, \cdot \rangle$, but usually refer to it as just $F(x)$.

We see that the operations are well-defined because the polynomials are well-defined and integral domain and therefore closed under addition and multiplication and do not have any zero divisors. Thus if neither $q(x)$ nor $g(x)$ are the zero polynomial, then neither is their product. This leads us to our next result that is analogous to the rational numbers.

Theorem 7.19. *Let F be a field. Then $\langle F(x), +, \cdot \rangle$ is a field.*

Proof. We leave the proof as an exercise to help the reader become more familiar with this field. □

This implies that the rational expressions with real valued coefficients, $\mathbb{R}(x)$ forms a field very similar to the rational numbers. We will explore this field in greater depth in Section 8.7.

7.4.1 Exercises

1. Prove that $\mathbb{R}(x)$ is a field.

Chapter 8

Real Valued Functions

This chapter builds upon the definitions of functions studied in Chapter 5, using the algebra material from Chapters 6 and 7, to study the interplay between the verbal, algebraic, tabular, and graphical representations of functions.

Related Content Standards

- (8.F.2) Compare properties of two functions each represented in a different way (algebraically, graphically, numerically in tables, or by verbal descriptions). *For example, given a linear function represented by a table of values and a linear function represented by an algebraic expression, determine which function has the greater rate of change.*
- (HSF.IF.9) Compare properties of two functions each represented in a different way (algebraically, graphically, numerically in tables, or by verbal descriptions).

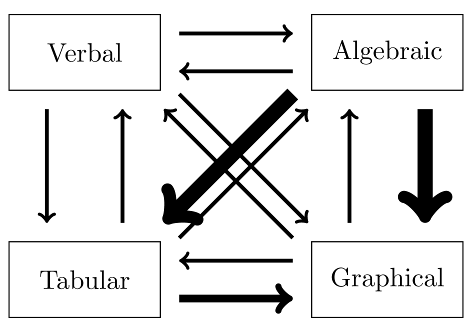


Figure 8.1: Emphases of translations between function representations

This process of translating between the different representations of a function is a key aspect of problem solving because different problems are more easily solved using different representations. For instance, if a set of data is given in a tabular format one would often use regression techniques to generate a graphical and algebraic representations of a function that models that data in order to interpolate or extrapolate information about the data within the context of the verbal situation given. Each of these different representations work together to help generate and interpret possible solutions to the problems determined. As such, it is important for students to be able to move from one representation to another.

In the secondary curriculum, students frequently move from algebraic expressions to graphical representations, either by hand or using technology, as stated in the Common Core Standard HSF.IF.7. However, students rarely translate from a graphical representation to an algebraic representation. We represent the different emphases between the translations by different sized arrows in Figure 8.1.

Related Content Standards

- (HSF.IF.7) Graph functions expressed symbolically and show key features of the graph, by hand in simple cases and using technology for more complicated cases.
  a. Graph linear and quadratic functions and show intercepts, maxima, and minima.
  b. Graph square root, cube root, and piecewise-defined functions, including step functions and absolute value functions.
  c. Graph polynomial functions, identifying zeros when suitable factorizations are available, and showing end behavior.
  d. Graph rational functions, identifying zeros and asymptotes when suitable factorizations are available, and showing end behavior.
  e. Graph exponential and logarithmic functions, showing intercepts and end behavior, and trigonometric functions, showing period, midline, and amplitude.

8.1 Function Properties

Throughout this chapter we will work to better understand the common functions in the secondary curriculum by studying certain properties of the functions.

Related Content Standards

- (8.F.5) Describe qualitatively the functional relationship between two quantities by analyzing a graph (e.g., where the function is increasing or decreasing, linear or nonlinear). Sketch a graph that exhibits the qualitative features of a function that has been described verbally.
- (HSF.IF.4) For a function that models a relationship between two quantities, interpret key features of graphs and tables in terms of the quantities, and sketch graphs showing key features given a verbal description of the relationship. Key features include: intercepts; intervals where the function is increasing, decreasing, positive, or negative; relative maximums and minimums; symmetries; end behavior; and periodicity.

8.1.1 Domain and Range

In this chapter we will deal with functions whose domain is a subset of the real numbers, $\mathbb{R}$. That domain may be inherited from the situation the function is modeling. For instance, a function that gives the cost of a purchase that is based on the number of items purchased would have a domain of the natural numbers. Where a function that models the number of feet above the ground that an object is located t seconds after being dropped would have a domain of the non-negative real numbers.

Frequently, a function is only defined by an algebraic representation. In this case, the implied domain for the function is the largest subset of $\mathbb{R}$ for which the function is defined. For example, the function $f(x) = \sqrt{x}$ is only defined for non-negative real numbers and so that would be the implied domain for the function.

Related Content Standards

- (HSF.IF.5) Relate the domain of a function to its graph and, where applicable, to the quantitative relationship it describes. For example, if the function $h(n)$ gives the number of person-hours it takes to assemble n engines in a factory, then the positive integers would be an appropriate domain for the function.

Once the domain for a function is determined, it is important to understand all of the possible elements in the co-domain that correspond to elements in the domain with the function, the range. The process for determining the range often involves many different techniques. If the original function is an injection, then one may be able to use the algebraic representation of the inverse function to determine the domain of the inverse function, which is the range of the original function. One may also use other properties of the function, such as its extreme points and its intervals of the domain on which the function is increasing or decreasing.

When graphing a function, we usually have the horizontal axis represent the domain of the function and the vertical axis represent the co-domain.

8.1.2 Maxima (minima) or relative maxima (relative minima)

It is often valuable to find the largest or smallest values that can be obtained by the function. These appear most obviously in optimization problems in calculus, where one finds the greatest (or least) values of f, or finds the greatest (or least) values on a given interval. These extreme values often depend upon the subset of the domain considered. When looking for a maximum or minimum value we find that for differentiable functions that these extreme values occur when the derivative is zero or undefined or at the endpoints of intervals in the domain considered. Other methods for finding these extreme values include using graphing calculators to graph the functions and have the technology determine these extreme values.

8.1.3 Increasing or decreasing

It is also valuable to determine those intervals on which a function, f, is increasing and those intervals on which f is decreasing. A function f is **increasing** on an interval (a, b) if for every $a < x \leq y < b$, $f(x) \leq f(y)$. The function f is **decreasing** on (a, b) if $a < x \leq y < b$ implies that $f(x) \geq f(y)$. We call the function strictly increasing or decreasing if the $\geq$ and $\leq$ symbols are replaced by strict inequality. If a function is either increasing or decreasing on its entire domain, then we call the function **monotonic**. One can also see that a strictly monotonic function is an injection and so we can find an inverse function when restricting the co-domain to the range.

8.1.4 Intercepts

When a function is providing a relationship between two variables, it is often useful to find the points at which either of the variables is zero. This process helps with understanding the graph of a function and often provides insight into the interpretation of the situation that a function may be modeling. While we often say that we are looking for the x-intercept and y-intercept of a function, it is important to have the flexibility that the variables have different names, especially in modeling situations. So we will try to use the convention of horizontal and vertical intercepts.

8.1.5 End behavior, singularities, and asymptotes

It is often helpful to understand what happens to the value of functions near the endpoints of a domain when the domain includes open intervals. For functions whose domain includes intervals of the form $(-\infty, c)$, (c, ∞), and $(-\infty, \infty)$, this is understanding what happens to $f(x)$ as $x \to \infty$ and/or as $x \to -\infty$. The study of this often focuses on horizontal asymptotes, but we will expand that perspective to a comparison of the behavior of functions 'near' $\pm\infty$ to the behavior of know functions. For instance, does the function grow at a similar rate to a certain polynomial function, a certain exponential function, or possibly a certain logarithmic function?

If the domain of the function includes a bounded open interval, it is important that we understand what happens to the output of the function as the input approaches the endpoints of the open interval. For

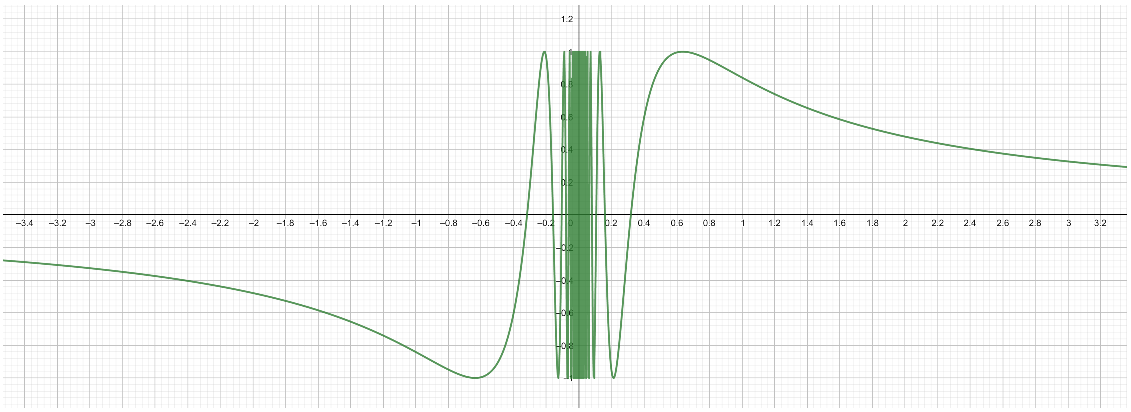


Figure 8.2: Graph of $y = \sin(1/x)$ ([Geogebra Application](https://www.geogebra.org/calculator/cd45ryxz))

instance for $f(x) = \frac{1}{x}$, as $x \to 0^+$ (x approaches zero from the right) $f(x) \to +\infty$. While as $x \to 0^-$, $f(x) \to -\infty$.

The function $y = f(x) = \sin\left(\frac{1}{x}\right)$ defined on the interval $(0, \infty)$ has a very unusual behavior near $x = 0$. The function oscillates between -1 and 1 an infinite number of times as x approaches 0.

8.1.6 Special Properties

In addition to the properties listed above, it is important to look for other properties of the function.

Is f a composition of known functions?

Consider $f(x) = \sqrt{-(x-1)(x+2)}$. This function is a composition of two functions that we already know and we can use knowledge about those functions to better understand this function. We will explore this process in Section 8.9.

Does the graph of f have a particular known shape?

When modeling a set of bivariate data, it is helpful to know if the graph of the data resembles the graph of a know function, i.e. linear, exponential, or quadratic. We can then use that information to determine the best options for choosing a function to model the situation.

Does the graph have symmetry?

If the graph is symmetric about a vertical line, then it is a transformation of an even function, and if it is symmetric about a point, it is a transformation of an odd function.

Definition 8.1. A function $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is **even** if $f(x) = f(-x)$ for all $x \in A$. The function is **odd** if $f(x) = -f(-x)$ for all $x \in A$.

Is f periodic?

It is also helpful to know if the function repeats itself over the domain. For instance, if someone wants to model the length of daylight for the day then it would be helpful to know that this is a periodic relationship that repeats itself every year.

Does f have and inverse function?

We often find it useful to 'undo' functions by using their inverse functions. When considering real-valued functions, this inverse function represents the same relationship between two variables, but viewed from a different perspective. Understanding this perspective also helps us to better understand the original function and to expand our ability to use such a function.

As we saw in Section 5.3, if a function is an injection it can be considered a bijection from its domain onto its range. For certain functions we may need to restrict the domain of the function in order to find an inverse.

Related Content Standards

- (HSF.BF.4) Find inverse functions.

  a. Solve an equation of the form $f(x) = c$ for a simple function f that has an inverse and write an expression for the inverse.
  b. Verify by composition that one function is the inverse of another.
  c. Read values of an inverse function from a graph or a table, given that the function has an inverse.
  d. Produce an invertible function from a non-invertible function by restricting the domain.

8.1.7 Exercises

1. For each of the following functions defined by their algebraic, verbal, tabular and/or graphical representation, describe the function according the the various properties described in this section. Furthermore, describe how these properties relate to the situation being modeled.

   a. A factory is built to produce chairs and for each day of operation it has a fixed cost of \$2,000 per day for rent, salaries, and other expenses, and a cost of \$80 per chair built for materials. We let $C(n)$ be the function that represents the cost of building n chairs in a day.

   b. A person is walking down a street at a constant rate. Let $D(t)$ be the distance from a certain lamppost t seconds from when the person starts walking down the street.

   c. Let f be the days-in-the-month function in a non-leap year defined by the table

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

   d. Let $G(L)$ be the grade points for a letter grade L given by

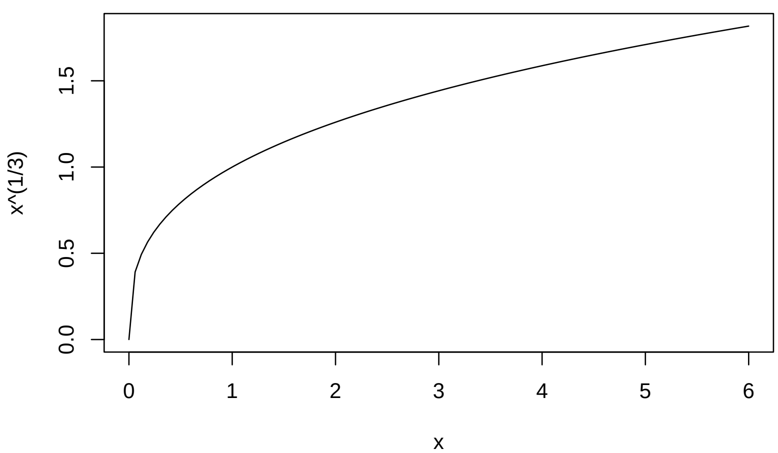
Letter grade, L	A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
Grade Points, G	4.33	4	3.67	3.33	3	2.67	2.33	2	1.67	1.33	1	.67	0

   e. The closing values of the S & P 500 Index each weekday for January 2000-December 2019. \begin{figure}

{

}
\end{figure}

a. Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \sqrt[3]{x}$, whose graph is given by \begin{figure}



{

}
\end{figure}

8.2 Transformations of Functions

When modeling using a function it is usually helpful to transform the function by a combination translations, dilations, and reflections in order to have the function fit the situation being modeled.

As an example we will look at a model of the number of hours of daylight in London, England for each day during 2020. We notice that the hours of daylight are periodic and that they resemble a trigonometric function. This function has a period of 365.25 days and we calculate the average length of daylight to be 12 hours and 14 minutes, which first occurs on the 81st day of the year. So we can model the situation with the function

$$f(x) = \left(\frac{\text{max hours of daylight} - \text{min hours of daylight}}{2}\right) \cdot \sin\left(\frac{2\pi}{365.25}(x-81)\right) + 12:14$$

As we can see in below, this function models the situation fairly well.

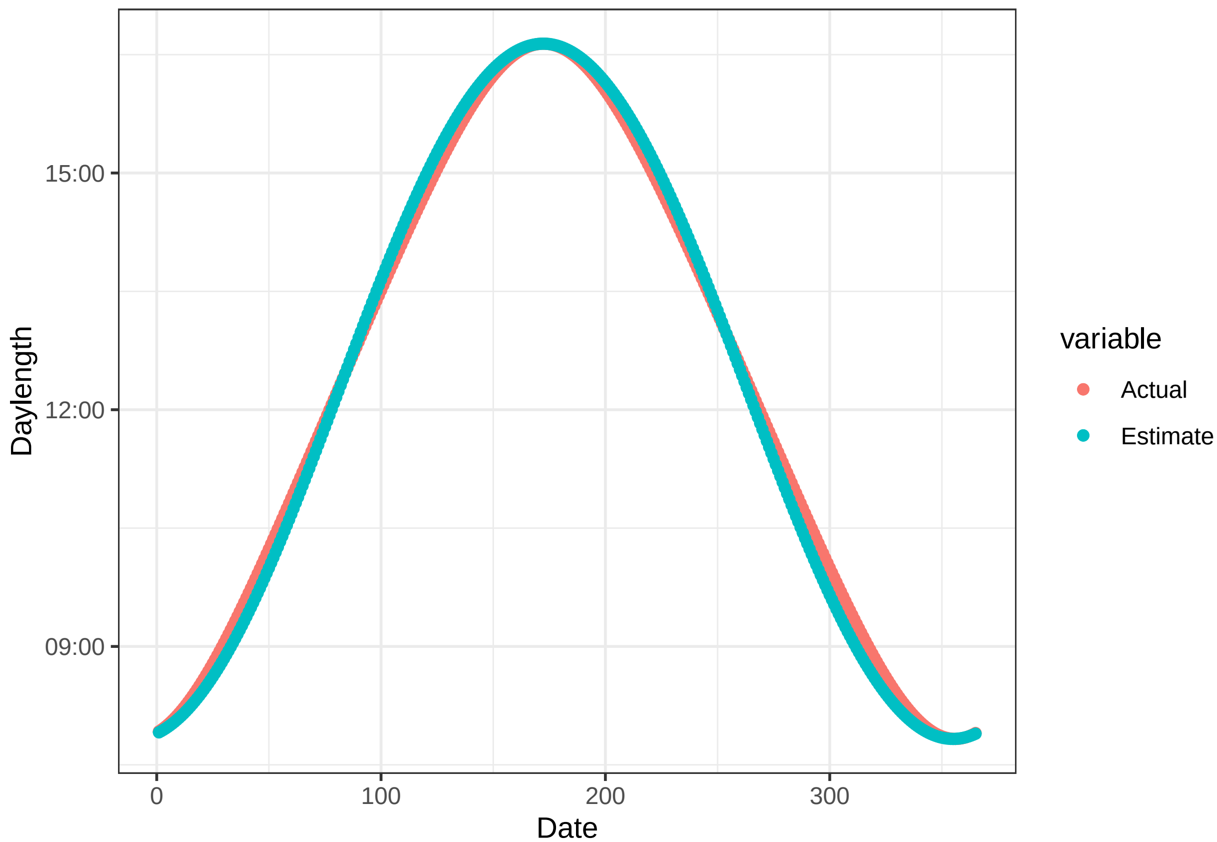


Figure 8.3: Hours of Daylight for London, 2020

We will study the family of functions generated from a parent function using these transformations throughout the rest of the chapter, using the function f given by the graph in Figure 8.4 as an example.

8.2.1 Horizontal Translations

Given a function f and a constant h, a new function $g(x) = f(x-h)$ is a **horizontal translation** of the function f. If h is positive, then $y = g(x)$ is a translation of $y = f(x)$ by h units to the right. If h is negative, it is a translation to the left.

Related Content Standards

- (HSF.BF.3) Identify the effect on the graph of replacing $f(x)$ by $f(x)+k$, $kf(x)$, $f(kx)$, and $f(x+k)$ for specific values of k (both positive and negative); find the value of k given the graphs. Experiment with cases and illustrate an explanation of the effects on the graph using technology. Include recognizing even and odd functions from their graphs and algebraic expressions for them.

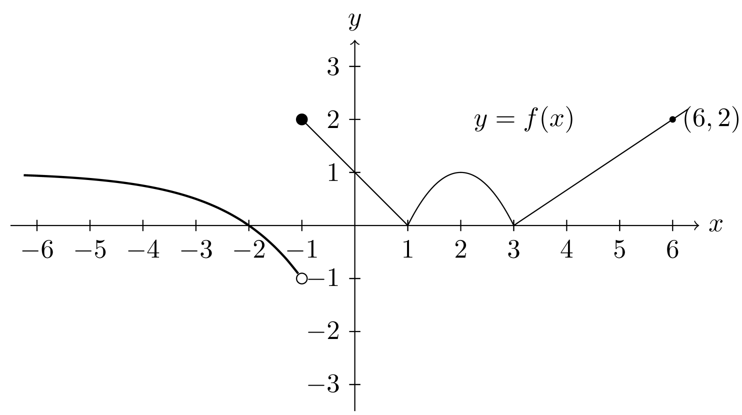


Figure 8.4: Tranformations: Base Function

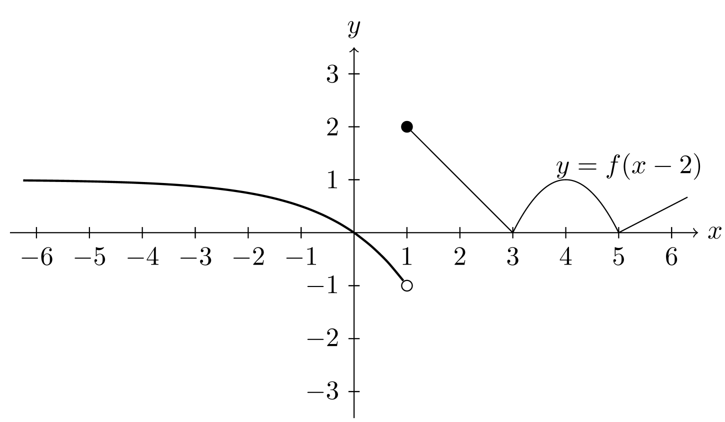


Figure 8.5: Tranformations: Horizontal Translation

8.2.2 Vertical Translations

Given a function f and a constant k, a new function $g(x) = f(x) + k$ is a **vertical translation** of the function f. If k is positive, then $y = g(x)$ is a translation of $y = f(x)$ by k units up. If k is negative, it is a translation down.

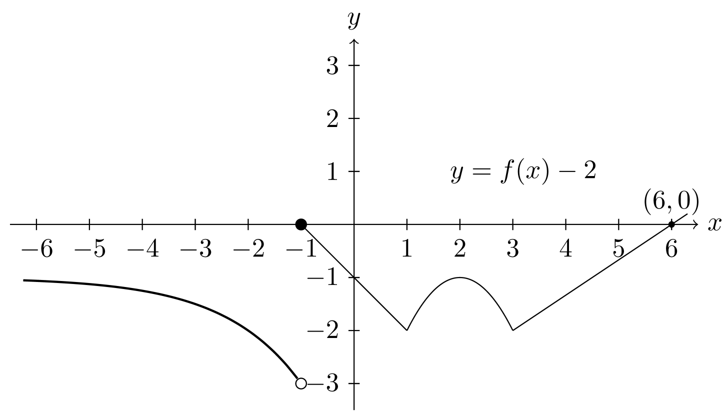


Figure 8.6: Tranformations: Vertical Translation

8.2.3 Horizontal Dilations

In addition to the translations, we can dilate the function away from and closer to the vertical axis. Given a function f and a constant $b > 0$, a new function $g(x) = f(bx)$ is a **horizontal dilation** by a factor of b. If $b > 1$, the new function is a compression towards the vertical axis by a factor of b. If $0 < b < 1$, the new function is a dilation away from the vertical axis by a factor of $\frac{1}{b}$.

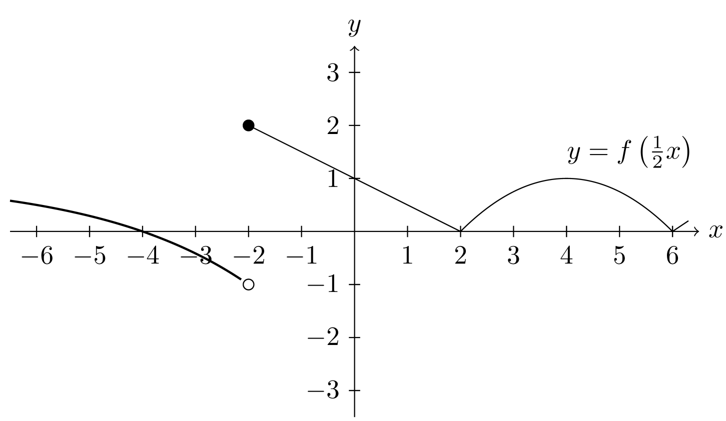


Figure 8.7: Tranformations: Horizontal Dilation

8.2.4 Vertical Dilations

Similarly to the horizontal dilations, multiplying the output of a function by a positive constant dilates the function vertically. Given a function f and a constant $a > 0$, the function $g(x) = af(x)$ is a dilation towards the horizontal axis if $0 < a < 1$ and away from the horizontal axis if $a > 1$.

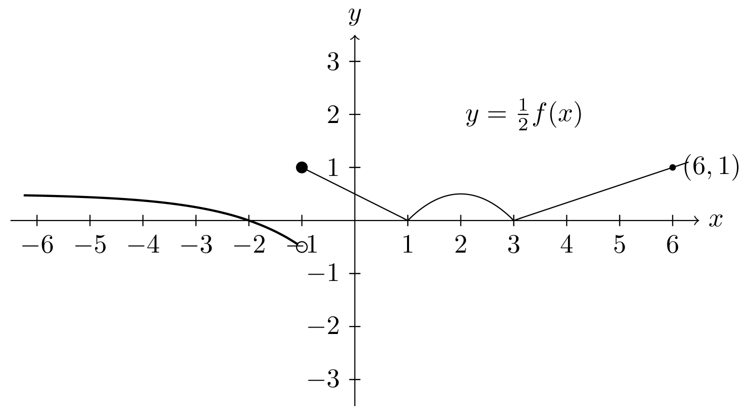


Figure 8.8: Tranformations: Vertical Dilation

8.2.5 Reflections

We are also able to reflect the graph of a function over the horizontal axis by changing the sign of the output of the function. So $g(x) = -f(x)$ is a **vertical reflection** of f about the horizontal axis.

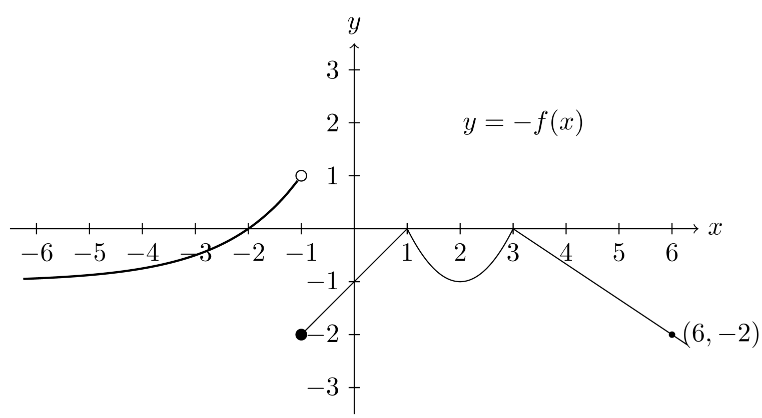


Figure 8.9: Tranformations: Vertical Reflection

By changing the sign of the variable in the domain, we can also reflect a function about the vertical axis so that the graph of $g(x) = f(-x)$ is a **horizontal reflection** of the graph of f about the vertical axis.

8.2.6 Families of Functions

One can combine all of these transformations together in order to generate a family of functions related to the original function f,

$$g(x) = a \cdot f(b(x-h)) + k.$$

In each of the next few sections we will study the family of functions related to the parent functions of $f(x) = x$, $f(x) = e^x$, $f(x) = \frac{1}{x}$, $f(x) = x^2$, and the basic trigonometric functions, along with each of their inverse functions when those functions exist.

8.2.7 Exercises

1. Let $f : [-3, 3] \to \mathbb{R}$ be given by the graph

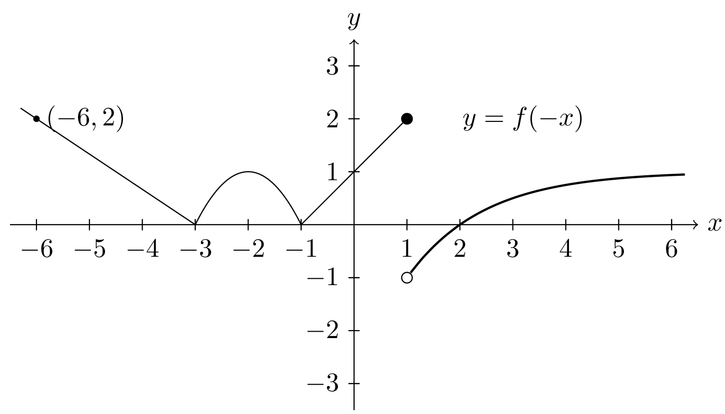
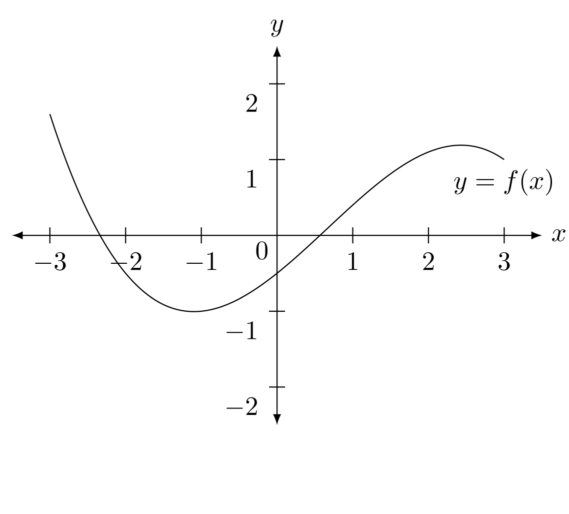


Figure 8.10: Tranformations: Horizontal Reflection



a. Sketch the graph of $y = f(2x)$.
b. Sketch the graph of $y = 2f(x)$.
c. Sketch the graph of $y = f(x+2)$.
d. Sketch the graph of $y = f(x) + 2$.
e. Skecth the graph of $y = -f(-x)$.

8.3 Linear and Exponential Functions

8.3.1 Linear Functions

The first parent function that we will study is the identity function $f(x) = x$. With this function, we see that horizontal and vertical dilations are connected in that $f(ax) = af(x) = ax$, and that these function describe proportional relationships between two variables. These proportional relationships are part of the core of the middle school mathematics curriculum, along with much of the physical science curriculum.

Related Content Standards

- (6.RP.1) Understand the concept of a ratio and use ratio language to describe a ratio relationship between two quantities.
- (6.RP.3) Use ratio and rate reasoning to solve real-world and mathematical problems.
- (7.RP.2) Recognize and represent proportional relationships between quantities.
- (8.EE.5) Graph proportional relationships, interpreting the unit rate as the slope of the graph.

While a direct proportional relationship fits some situations very well, such as Hooke's law that the force needed to compress a spring by a distance is directly proportional to that distance, many others can be modeled by a proportional relationship near certain values. In fact, one of the foundations of calculus is that most functions that we use in modeling can locally be approximated by a proportional relationship. However, to model these proportional relationships we often need to translate the function to a point on the function. Such a translation gives us that for any 'differentiable' function, f, near a point $x = a$ we can approximate f by

$$f(x) \approx f'(a)(x-a) + f(a).$$

As such, we will study the family of linear functions of the form

$$f(x) = m(x-h) + k,$$

often described as the point-slope form of a line with slope m through the point (h, k). These functions can also be written in a slope intercept form, $f(x) = mx + b$ where $b = k - mh$.

If $f(x) = mx + b$, then we can see that the domain and range of the function is all of the real numbers. If $m > 0$ the function is strictly increasing, while if $m < 0$ it is strictly decreasing. As such, the function has no extrema.

We also know that since linear functions are monotonic, they are invertible. If $y = f(x) = mx + b$, the inverse of f is $x = f^{-1}(y) = \frac{1}{m}y - \frac{b}{m}$.

8.3.2 Exponential Functions

In Section 6.2 we found a strong relationship between linear functions and exponential functions with the additive relationship of linear functions corresponding with the multiplicative relationship of exponential functions. In the case of exponential functions we will have our parent function be the function $f(x) = e^x$, making the family of exponential functions to be written in the form

$$af(b(x-h)) + k = ae^{b(x-h)} + k = (ae^{-bh}) \cdot e^{bx} + k$$

which means that we can generate all possibilities with functions of the form $g(x) = ae^{bx} + k$

When introducing exponential functions in middle school, the base is usually either 2 or 10 since students have not yet been introduced to e, but we can see that each different base is actually a horizontal dilation of the function with base e.

We saw in Section 4.6 that if $b > 0$ and $x < y$, then $e^b > 1$ and $e^{bx} < e^{by}$, making e^{bx} a strictly increasing function. Similarly if $b < 0$, e^{bx} is a strictly decreasing function. So we have the following possible shapes for functions of the form $g(x) = ae^{bx}$.

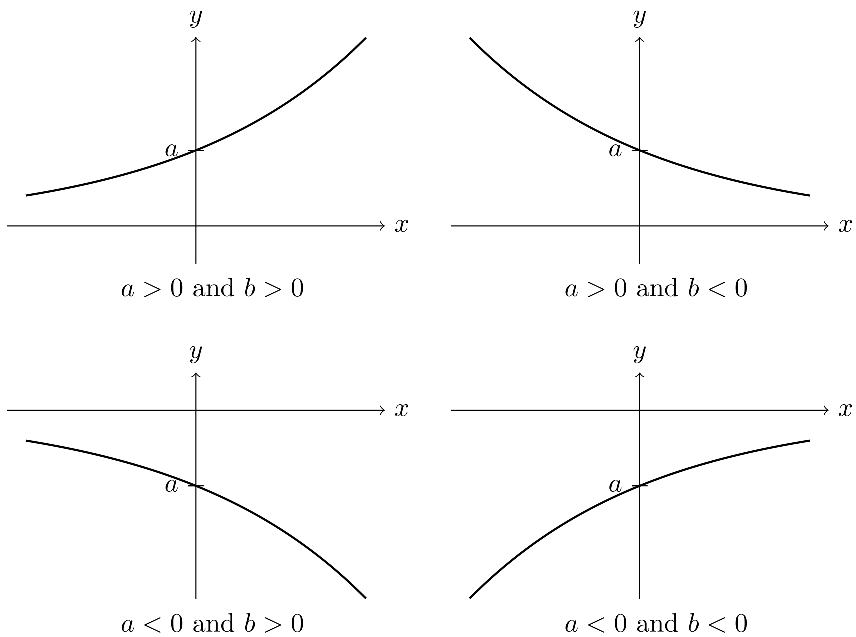


Figure 8.11: Graphs of Exponential Functions

We can see that with each of these cases that the function is monotonic and the range is either $(0, \infty)$ or $(-\infty, 0)$, depending on the sign of a. If we perform a vertical translation to acheive the entire family of exponential functions then we see that we have the following properties for $f(x) = ae^{bx} + k$:

- The domain is $\mathbb{R}$ and the range is (k, ∞) if $a > 0$ and $(-\infty, k)$ if $a < 0$.
- The function is strictly monotonic.
- If $b > 0$, then $e^b > 1$, $f(x) \to k$ as $x \to -\infty$, and $f(x) \to \pm\infty$ as $x \to \infty$ (depending on the sign of a). These functions model exponential growth.
- If $b < 0$, then $0 < e^b < 1$, $f(x) \to k$ as $x \to \infty$, and $f(x) \to \pm\infty$ as $x \to -\infty$ (depending on the sign of a). These functions model exponential decay.

Related Content Standards

- (HSF.IF.8) Write a function defined by an expression in different but equivalent forms to reveal and explain different properties of the function.

  b. Use the properties of exponents to interpret expressions for exponential functions. For example, identify percent rate of change in functions such as $y = (1.02)^t$, $y = (0.97)^t$, $y = (1.01)12^t$, $y = \frac{(1.2)^t}{10}$, and classify them as representing exponential growth or decay.

8.3.3 Logarithmic Functions

Since the exponential functions are monotonic, they are bijections onto their range. If we let $f(x) = e^x$, then we know that $f : \mathbb{R} \to (0, \infty)$ is a bijection. From Chapter 5 we know that there is a unique function $g : (0, \infty) \to \mathbb{R}$ that is also a bijection with $g(f(x)) = x$ for all $x \in \mathbb{R}$ and $f(g(x)) = x$ for all $x \in (0, \infty)$. We will call this function the natural logarithm and denote it by ln.

We can then use properties of the exponential function to determine properties of the natural logarithm.

If $\alpha, \beta > 0$, then there exist a and b so that $\alpha = e^a$ and $\beta = e^b$, and

$$\ln(\alpha\beta) = \ln\left(e^a \cdot e^b\right) = \ln\left(e^{a+b}\right) = a + b = \ln(\alpha) + \ln(\beta).$$

We also have that

$$\ln\left(\frac{\alpha}{\beta}\right) = \ln\left(\frac{e^a}{e^b}\right) = \ln\left(e^{a-b}\right) = a - b = \ln(\alpha) - \ln(\beta).$$

If $\alpha > 0$ and $x \in \mathbb{R}$, then there exists an $a \in \mathbb{R}$ such that $\alpha = e^a$. Then

$$\ln\left(\alpha^x\right) = \ln\left(\left(e^a\right)^x\right) = \ln\left(e^{ax}\right) = ax = x\ln(\alpha).$$

Related Content Standards

- (HSF.BF.5) Understand the inverse relationship between exponents and logarithms and use this relationship to solve problems involving logarithms and exponents.

Since a^x $(a > 0)$ has the same properties as e^x it also has an inverse function that we will call logarithm base a and denote it by

$$\log_a : (0, \infty) \to \mathbb{R}.$$

We can find the relationships between the various bases. If $y = \log_a(x)$, then $x = a^y$. We can evaluate both sides of the equation in the natural log function so that

$$\ln(x) = \ln\left(a^y\right) = y\ln(a)$$

and so

$$\log_a(x) = \frac{\ln(x)}{\ln(a)}.$$

Using these properties together we are able to use logarithm functions to find equivalent equations to equations such as $ab^{ct} = d$. By inputing both sides of the equation into the ln function, we have that

$$\ln(a) + (ct)\ln(b) = \ln(d).$$

We can then find the equivalent equation of

$$t = \frac{\ln(d) - \ln(a)}{c\ln(b)}.$$

Related Content Standards

- (HSF.LE.4) For exponential models, express as a logarithm the solution to $ab^{ct} = d$ where a, c, and d are numbers and the base b is 2, 10, or e; evaluate the logarithm using technology.

8.3.4 Relationships between Linear and Exponential

Since linear functions have a constant additive rate of change and exponential functions have a constant multiplicative rate of change, it is important to recognize the types of situations that are more likely to be modeled with a linear function and which are more likely modeled with an exponential function.

Related Content Standards

- (HSF.LE.1) Distinguish between situations that can be modeled with linear functions and with exponential functions.

  a. Prove that linear functions grow by equal differences over equal intervals, and that exponential functions grow by equal factors over equal intervals.
  b. Recognize situations in which one quantity changes at a constant rate per unit interval relative to another.
  c. Recognize situations in which a quantity grows or decays by a constant percent rate per unit interval relative to another.

Some of the factors that point towards a linear model are adding a fixed amount over fixed intervals in the domain, such as adding $200 to a savings account every month. Factors that direct the model to be exponential is that the amount of growth would depend upon the current value, such as interest rates or cell division in biology.

When given a set of data, the function used to model the data sometimes depend upon the question being asked about the data. Consider the following population data for the State of Alabama[1].

Year	1800	1850	1900	1950	2000
Population	$1,250$	$771,623$	$1,828,697$	$3,061,743$	$4,447,100$

This data can be modeled using either a linear regression or an exponential regression. However, if you include the population in 1800 in the exponential regression, it estimates the population in 2050 to be nearly 100 million people (a highly unlikely number). If one only uses the data points from 1900, 1950, and 2000, the exponential model predicts a population of around 7 million in 2050. A linear model with all of the data points estimates the population in 2050 to be around 5.3 million people (a much more likely scenario).

If, however, you are wanting to estimate the population in years prior to 1800, the linear model becomes useless as it predicts a negative population. So for this type of question, an exponential function would be more useful.

8.3.5 Exercises

1. The salary scales in three school districts are as follows, for a teacher with a master's degree:

   - District P: $30,000 plus $1500 for each year of experience
   - District Q: $30,000 plus $1750 for each year of experience
   - District R: $28,000 plus $1750 for each year of experience

   a. Give a formula for the salary in each district for a teacher with n years of experience at the start of the school year.
   b. Use your formulas to indicate the number of years experience teachers in Districts P and R when they earn the same salary.
   c. Use your formulas to indicate the number of years experience teachers in Districts P and Q when they earn the same salary.

[1] https://en.wikipedia.org/wiki/Demographics/_of/_Alabama, retrieved 6/25/2020

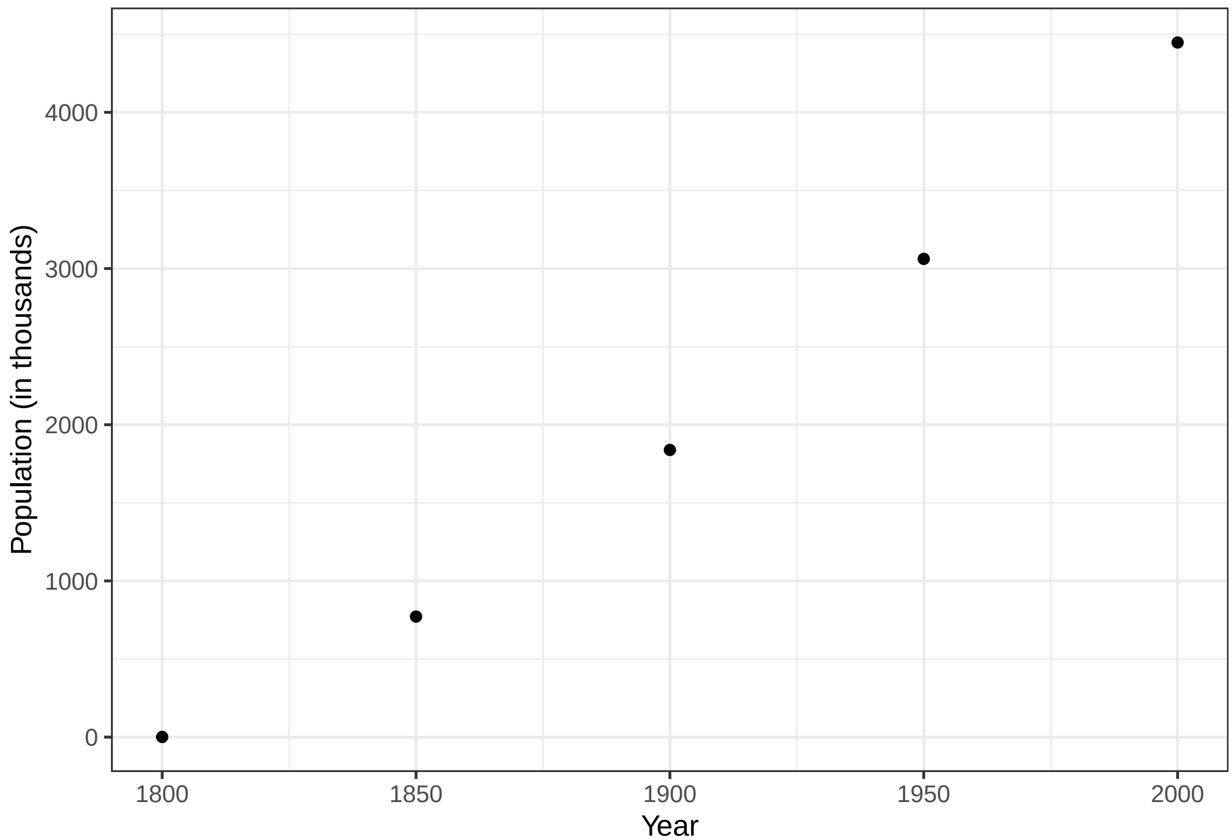


Figure 8.12: Population of Alabama

Table 8.1: Race Times for Olympic Men's 100m Dash

Year	Time (s)	Runner	Country
1948	10.30	Harrison Dillard	USA
1952	10.40	Lindy Remingino	USA
1956	10.50	Bobby Morrow	USA
1960	10.20	Armin Hary	Germany
1964	10.00	Bob Hayes	USA
1968	9.95	Jim Hines	USA
1972	10.14	Valery Borzov	USSR
1976	10.06	Hasely Crawford	Trinidad
1980	10.25	Allan Wells	Great Britain
1984	9.99	Carl Lewis	USA
1988	9.92	Carl Lewis	USA
1992	9.96	Linford Christie	Great Britain
1996	9.84	Donovan Bailey	Canada
2000	9.87	Maurice Greene	USA
2004	9.85	Justin Gatlin	USA
2008	9.69	Usain Bolt	Jamaica
2012	9.63	Usain Bolt	Jamaica
2016	9.81	Usain Bolt	Jamaica
2021	9.80	Marcell Jacobs	Italy

d. Use your formulas to indicate the number of years experience teachers in Districts Q and R when they earn the same salary.
e. If in District T, teachers earn a salary S_T dollars plus E_T dollars for each year of experience, and in District U, teachers earn a salary of S_U dollars plus E_U dollars for each year of experience, $S_T > S_U$, and $E_U > E_T$, how many years will it take District U teachers to catch up to District T?
f. For each of the previous scenarios, at what value of n will teachers in the two district have earned the same amount of total money during the previous n years?

2. The following are winning times for the Men's 100m dash at the Olympics.

   Find a mathematical function that model this behavior and use this function to predict the winning time in 2060. Give reasoning for your work, including the assumptions made for your models.

3. The number of bacteria in a certain population increases according to a continuous exponential growth model, with a growth rate parameter of 3.25% per hour. How many hours does it take for the size of the sample to double?

4. LaVonda is choosing between 2 jobs. For Job A, she would earn $\$27,000$ the first year and each year after that she would get a raise of $\$3,000$. For Job B, she would earn $\$30,000$ the first year and each year after that she would get a raise of 4% of the previous year's salary.

   a. What is the first year during which LaVonda's salary for Job A would exceed that of Job B? Show your work, and explain the approach you used to find your answer.
   b. Which year would the total amount earned since starting Job A first exceed the total amount earned since starting Job B? Show your work, and explain the approach you used to find your answer.
   c. Using sequence and series formulas, what would LaVonda's yearly salary and total amount of money earned be after 20 years at Job A? What would her yearly salary and total amount of money earned be after 20 years at Job B?

5. A business owner spent \$500 on start-up fees to produce and sell candles. Each candle costs an additional \$3.00 to produce. What is the minimum number of candles that the owner must produce for the average cost per candle to be less than \$3.75?

8.4 Linear Fractional Transformations

The function $f(x) = \frac{1}{x}$ is another function that appears in physical applications such as Boyle's law that the pressure of a given mass of an ideal gas is inversely proportional to its volume at a fixed temperature. In this section we will explore the family of functions that derive from this function. If $f(x) = \frac{1}{x}$, then we see that for $\alpha \neq 0$ and $\beta \neq 0$

$$g(x) = \alpha \cdot f(\beta(x-h)) + k = \frac{\alpha}{\beta} \cdot \frac{1}{x-h} + k$$

which can be rewritten as

$$g(x) = \frac{\alpha + k(\beta x - \beta h)}{\beta x - \beta h} = \frac{(k\beta)x + (\alpha - \beta hk)}{\beta x - (\beta h)} = \frac{kx + \left(\frac{\alpha}{\beta} - hk\right)}{x-h}.$$

On the other hand, any function of the form

$$h(x) = \frac{ax+b}{cx+d}$$

can be rewritten in the form

$$h(x) = \frac{a}{c} - \left(\frac{ad-bc}{c^2}\right)\frac{1}{x+\frac{d}{c}}$$

and so we see that it is a transformation of the inverse proportional function if, and only if, $ad - bc \neq 0$ and $c \neq 0$.

If we allow $c = 0$, then the functions are transformations of both the linear and inversely proportional parent functions.

Therefore, the family of functions whose parent function is either $f(x) = \frac{1}{x}$ or $f(x) = x$ we call linear fractional transformations (also called a M"{o}bius transformations) as a real-valued function of the form

$$f(x) = \frac{ax+b}{cx+d}$$

where a, b, c, and d are real-valued constants with $(ad - bc) \neq 0$.

8.4.1 Domain

If $c \neq 0$, then these functions are not defined at the point $\frac{-d}{c}$ and so the natural domain for the function would be $\mathbb{R} \setminus \{\frac{-d}{c}\}$.

If $c = 0$, then $d \neq 0$ and so we see that the function becomes a linear function of the form

$$f(x) = \frac{a}{d}x + \frac{b}{d},$$

and so the domain of the function in $\mathbb{R}$.

8.4.2 Range

If $c \neq 0$, the parent function for these functions is inversely proportional function so that

$$f(x) = \frac{ax+b}{cx+d} = \frac{a}{c} - \left(\frac{ad-bc}{c^2}\right)\frac{1}{x+\frac{d}{c}}$$

So for this case that $c \neq 0$, we see that this function is a transformation of the parent function that takes transforms the plane so that the horizontal axis maps to the horizontal line of $y = \frac{a}{c}$, and so the range of the function is $\mathbb{R} \setminus \{\frac{a}{c}\}$.

If $c = 0$ the linear fractional transformation reduces to a linear function whose range is $\mathbb{R}$.

8.4.3 Increasing or decreasing

Using the graph of $y = \frac{1}{x}$ we can sketch a graph of the transformation of this function.

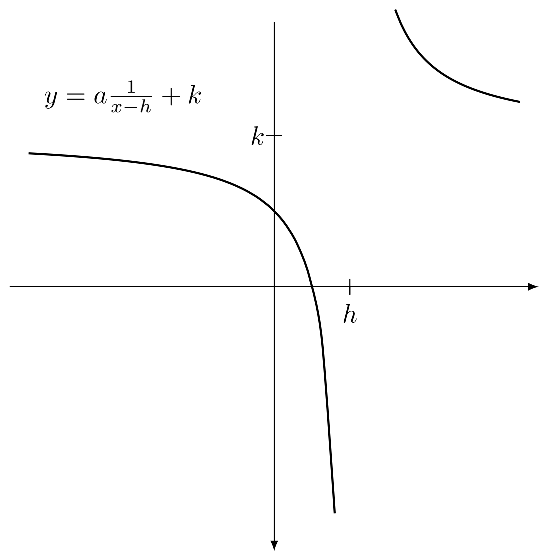


Figure 8.13: Transformations of $f(x) = 1/x$

By using the expression of the linear fractional transformation of

$$f(x) = \frac{ax+b}{cx+d} = \frac{a}{c} - \left(\frac{ad-bc}{c^2}\right)\frac{1}{x+\frac{d}{c}}$$

we see that if $ad - bc > 0$ that f is increasing on $(-\infty, \frac{-d}{c})$ and on $(\frac{-d}{c}, \infty)$. We also see that if $ad - bc < 0$ that the function is decreasing on these intervals.

8.4.4 Intercepts

From the primary algebraic expression of the function one can see that if $a \neq 0$ the horizontal intercept of the function is at $\frac{-b}{a}$, and if $a = 0$ that there is no horizontal intercept.

One can also see that if $d \neq 0$ that the vertical intercept is at $\frac{b}{d}$ with no vertical intercept if $d = 0$.

8.4.5 End behavior and asymptotes

One can see from the graph of $f(x) = \frac{ax+b}{cx+d} = \frac{a}{c} - \left(\frac{ad-bc}{c^2}\right)\frac{1}{x+\frac{d}{c}}$ that

$$\lim_{x\to\pm\infty} f(x) = \frac{a}{c}$$

and that the behavior near the vertical asymptote of $x = -\frac{d}{c}$ depends upon the sign of $ad - bc$.

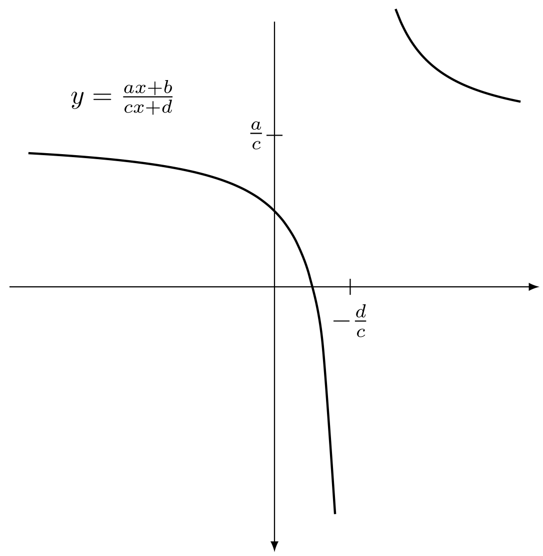


Figure 8.14: General Linear Fractional Transformation

8.4.6 Extensions

By adding a single element to the set of real numbers that we will call ∞, we can redefine these functions to have both a domain and range of $\mathbb{R} \cup \{\infty\}$. If $c \neq 0$ we define f on this extended domain to be

$$f(x) = \begin{cases} \frac{ax+b}{cx+d}, & x \in \mathbb{R} \setminus \{-\frac{d}{c}\} \\ \frac{a}{c}, & x = \infty \\ \infty, & x = -\frac{d}{c} \end{cases}$$

and if $c = 0$,

$$f(x) = \begin{cases} \frac{a}{d}x + \frac{b}{d}, & x \in \mathbb{R} \\ \infty, & x = \infty \end{cases}.$$

For the ease of notation, we will denote these extended functions by the algebraic representation of the original function and imply the extension to the extended real numbers.

From the results above about range and monotonicity of the function we can see that the linear fractional transformations on these extended domains and co-domains are bijections. That leads us to want to understand the forms of the inverse function and the composition of two of these linear fractional transformations.

In order to understand more about the composition of two linear fractional transformations we will define

$$f(x) = \frac{ax+b}{cx+d} \text{ and } g(x) = \frac{\alpha x+\beta}{\gamma x+\delta}.$$

Then we have that

$$\begin{aligned}(g \circ f)(x) &= \frac{\alpha\left(\frac{ax+b}{cx+d}\right)+\beta}{\gamma\left(\frac{ax+b}{cx+d}\right)+\delta} \\ &= \frac{\alpha ax+\alpha b+\beta cx+\beta d}{\gamma ax+\gamma b+\delta cx+\delta d} \\ &= \frac{(\alpha a+\beta c)\,x+(\alpha b+\beta d)}{(\gamma a+\delta c)\,x+(\gamma b+\delta d)}\end{aligned}$$

Since the composition of two bijections is again a bijection, this composition of two linear fractional transformations is again a linear fractional transformation.

In order to determine the form of the inverse function of a linear fractional transformation, $f(x) = \frac{ax+b}{cx+d}$ we will use the form of the composition of two linear fractional transformations given above and let

$$(\alpha b+\beta d) = 0,\ (\gamma a+\delta c) = 0, \text{ and}$$

$$(\alpha a+\beta c) = (\gamma b+\delta d)$$

so that the composition becomes the identity function.

If $c = 0$, then

$$f(x) = \frac{a}{d}x + \frac{b}{d}$$

and the second equation gives us that $\gamma = 0$, and then using the third equation we have that $\alpha a = \delta d$. This then implies that

$$f^{-1}(x) = \frac{d}{a}x + \frac{\beta}{\delta}$$

and since $\beta = \frac{-\alpha b}{d}$ we see that $\frac{\beta}{\delta} = -\frac{b}{a}$ and so

$$f^{-1}(x) = \frac{d}{a}x - \frac{b}{a}.$$

If $c \neq 0$, then $\delta = -\frac{\gamma a}{c}$ and $\beta = \frac{-\alpha ac+\gamma bc+\gamma ad}{c^2}$.

Another method to determine the form of the inverse function is based on the graph of the original function. Since, if $c \neq 0$,

$$f(x) = \frac{ax+b}{cx+d} = \frac{a}{c} - \left(\frac{ad-bc}{c^2}\right)\frac{1}{x+\frac{d}{c}}$$

we see that f is undefined for $x = \frac{-d}{c}$ giving that f^{-1} will not have $\frac{-d}{c}$ in its range. Similarly, since $\frac{a}{c}$ is not in the range of f, it is not in the domain of f^{-1}. So if f^{-1} is a linear fractional transformation it is of the form

$$f^{-1}(x) = \frac{-d}{c} + \omega\frac{1}{x-\frac{a}{c}} = \frac{-dx+\frac{ad}{c}+c\omega}{cx-a} = \frac{-dx+\beta}{cx-a}$$

for some $\omega \in \mathbb{R}$ and where $\beta \in \mathbb{R}$ is dependent upon this ω. We can then use the algebraic representation of the composition of functions to see that $-db+\beta d = 0$ showing that $\beta = b$ and so

$$f^{-1}(x) = \frac{-dx+b}{cx-a}.$$

A third method of determining the inverse function of

$$f(x) = \frac{ax+b}{cx+d}$$

when $c \neq 0$ is to rewrite the equation

$$y = \frac{ax+b}{cx+d}$$

so that we have

$$x = \frac{dy-b}{-cy+a}$$

giving

$$f^{-1}(x) = \frac{dx-b}{-cx+a}.$$

With each of these methods, we see that the inverse function of a linear fractional transformation is another linear fractional transformation.

8.4.7 Exercises

1. Prove that the set of linear fractional transformations, together with the operation of function composition forms a group.
2. Is the group of linear fractional transformations with function composition abelian?
3. Compare and contrast the group of linear functional transformations under function composition and the group of invertible 2×2 matrices under matrix multiplication.

8.5 Quadratic Polynomials

Related Content Standards

- (HSF.IF.8) Write a function defined by an expression in different but equivalent forms to reveal and explain different properties of the function.
    a. Use the process of factoring and completing the square in a quadratic function to show zeros, extreme values, and symmetry of the graph, and interpret these in terms of a context.

Quadratic functions are polynomials whose highest non-zero coefficient is of degree 2 and are often written in the form

$$f(x) = a_0 + a_1 x + a_2 x^2.$$

In order to avoid the use of indices for students when they are working with these quadratic functions, we write them in the form

$$f(x) = ax^2 + bx + c.$$

8.5.1 Completing the Square

To better understand the graph of these quadratics we will modify this form of the quadratic function using a process called completing the square. In this process we rewrite the polynomial in order to find a monic polynomial without a constant coefficient,

$$f(x) = a\left(x^2 + \frac{b}{a}x\right) + c.$$

We now look at the area model for the polynomial $x^2 + \frac{b}{a}x$ by splitting the $\frac{b}{a}x$ into two pieces.

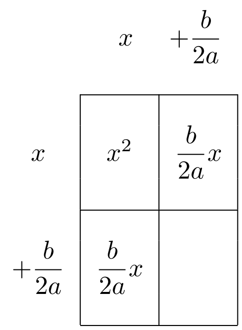


Figure 8.15: Completing the Square: Area Model

Then we see that the area needed to complete the square is $\frac{b^2}{4a^2}$ and so we adjust the polynomial accordingly to

$$f(x) = a\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} - \frac{b^2}{4a^2}\right) + c.$$

We are now able to distribute the a through the parenthesis to get

$$f(x) = a\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}\right) - \frac{b^2}{4a} + c$$

and we see that the expression inside of the parenthesis is the area of the square. Hence,

$$f(x) = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c.$$

Therefore, we can write all quadratic polynomials in the form

$$f(x) = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a}.$$

We can also look at quadratic polynomials from the perspective of transformations of the graph. Beginning with $f(x) = x^2$, we can see that the family of functions generated from this function are of the form

$$h(x) = a(x-h)^2 + k$$

which we will call the vertex form of the quadratic function. Equating this with the prior expression we see that all quadratic polynomials can be written in this way and that the vertex of a polynomial of the form

$$f(x) = ax^2 + bx + c$$

is at the point

$$\left(-\frac{b}{2a}, -\frac{b^2 - 4ac}{4a}\right).$$

We also see from properties of the transformations of the graph that if $a > 0$ the parabola opens up and if $a < 0$ the parabola opens down. Thus the range of the function $f(x) = ax^2 + bx + c$ is

$$\left[-\frac{b^2-4ac}{4a}, \infty\right) \text{ if } a > 0 \quad \text{and} \quad \left(-\infty, -\frac{b^2-4ac}{4a}\right] \text{ if } a < 0.$$

Related Content Standards

- (HSA.SSE.3) Choose and produce an equivalent form of an expression to reveal and explain properties of the quantity represented by the expression.
  b. Complete the square in a quadratic expression to reveal the maximum or minimum value of the function it defines.

8.5.2 Factoring

In addition to the standard form of $f(x) = ax^2 + bx + c$ and the vertex form of $f(x) = a(x-h)^2 + k$, it is useful to write quadratic functions in their factored form of $f(x) = (x - d_1)(x - d_2)$ as this form helps to determine the horizontal intercepts of the graph of the function.

To move from the standard form to the factored form, one uses the vertex form of

$$f(x) = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a}$$

and then find the values of x for which the function is equal to zero. These points are the values d such that

$$a\left(d + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a} = 0.$$

Using some algebraic properties we can rewrite this equation as

$$\left(d + \frac{b}{2a}\right)^2 = \frac{b^2 - 4ac}{4a^2}.$$

We then see that there are two values for d that make this equation true.

$$d_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad \text{and} \quad d_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

Therefore, the factored form of the quadratic can be written as

$$f(x) = \left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right)\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a}\right).$$

Related Content Standards

- (HSA.SSE.3) Choose and produce an equivalent form of an expression to reveal and explain properties of the quantity represented by the expression.
  a. Factor a quadratic expression to reveal the zeros of the function it defines.
- (HSN.CN.7) Solve quadratic equations with real coefficients that have complex solutions.

8.5.2.1 Polynomial Identities

While the quadratic formula works every time for factoring a quadratic polynomial, there are a couple polynomial identities that are very useful when factoring specific types of quadratics. The first is when we see a difference of squares.

$$x^2 - b^2 = (x - b)(x + b)$$

This difference of squares can also be used to factor $x^4 - 16$ as $(x^2 - 4)(x^2 + 4)$. We can then rewrite this to find more differences of squares so that

$$\begin{aligned} x^4 - 16 &= (x^2 - 4)(x^2 + 4) = (x^2 - 4)(x^2 - (-4)) \\ &= (x - 2)(x + 2)(x - 2i)(x + 2i). \end{aligned}$$

Another quadratic form that creates a simpler factorization is a perfect square.

$$x^2 + 2ax + a^2 = (x + a)^2$$

This can be recognized when the absolute value of the middle term of a quadratic is twice the product of the square roots of the other two terms. Some examples of uses include $4x^2 + 12x + 9$ or $12y^2 - 8\sqrt{3}y + 4$.

Related Content Standards

- (HSN.CN.8) Extend polynomial identities to the complex numbers. *For example, rewrite* $x^2 + 4$ *as* $(x + 2i)(x - 2i)$.

8.5.3 Inverse Function

We notice that a quadratic polynomial does not have an inverse. However, if we restrict the domain of the quadratic polynomial to an interval on which it is monotonic, we can find an inverse. So if we look at the function

$$f : [0, \infty) \to [0, \infty) \text{ with } f(x) = x^2$$

we see that f is a bijection. We define the inverse function as

$$g : [0, \infty) \to [0, \infty) \text{ with } g(x) = \sqrt{x}.$$

We will look at this process of restricting the domain in order to find the inverse function in more detail in the section on trigonometric functions.

8.5.4 Exercises

1. Let m and n be the zeros of the function $f(x) = ax^2 + bx + c$.

   a. Find a formula for $m + n$ in terms of a, b, and c.
   b. Find a formula for $m^2 + n^2$ in terms of a, b, and c.
   c. Find a formula for $m^3 + n^3$ in terms of a, b, and c.

2. Given $f(x) = -ax^2 + 2ahx - ah^2 + k$, find the domain and range of $f(x)$ where $a > 0$, $h > 0$, and $k > 0$.

3. Given $(k - 1)x^2 + kx + 1 = 0$, where -1 is one solution, what is the other solution?

4. Let $y = f(x)$ be a parabola with vertex $(4, 1)$ that passes through the point $(3, -7)$.

   a. What are the domain and range of $f(x)$? Explain how you found your answers based on just the vertex and the point given.

   b. What is the equation of this parabola in vertex form? Show your work algebraically. Explain how you used the points indicated on the graph to determine the equation.

5. Let $d(n)$ stand for the number of diagonals of a polygon of n sides. Here is a table of values of $d(n)$.

   a. Use the table of values to find a polynomial formula for $d(n)$ in terms of n.
   b. Give a geometric argument to show that your formula is true for all n.

6. The top and bottom margins of a poster are each 6 cm and the side margins are each 4 cm. If the area of the printed material on the poster is fixed at 384 cm^2, find the dimensions of the poster with the smallest area.

Table 8.2:

\$n\$	\$d(n)\$
3	0
4	2
5	5
6	9
7	14
8	20
9	27
10	35
11	44

8.6 General Polynomial Functions

Now that we understand the graphs of quadratic polynomials we turn our attention to polynomials of higher degree. For all polynomials, the domain is the set of all real numbers. The range, however, are much harder to define and are much easier to determine using numerical methods in computer software.

8.6.1 Short-Term Behavior

Related Content Standards

- (HSA.APR.3) Identify zeros of polynomials when suitable factorizations are available, and use the zeros to construct a rough graph of the function defined by the polynomial.

We know that the graph of a polynomial p intercepts the horizontal axis at every point $x = c$ such that $p(c) = 0$. We also know from the factor theorem that this corresponds to $x - c$ being a factor of $p(x)$. We say that p has a zero of order m at a point $x = c$ if the prime factorization of $p(x)$ includes $(x - c)^m$ as a factor, or equivalently

$$p(x) = (x - c)^m \cdot q(x)$$

where $q(x)$ is a polynomial for which $x - c$ is not a factor.

Theorem 8.1. *Let p be a polynomial with a zero of order m at $x = c$. Then*

$$p(x) = (x - c)^m \cdot q(x)$$

where $q(x)$ is a polynomial with $q(c) \neq 0$. Furthermore, the ***Taylor polynomial*** *of degree m for p centered at $x = c$ is*

$$T_m(x) = q(c) \cdot (x - c)^m.$$

Before we prove this theorem, we will prove the following lemma.

Lemma 8.1 (General Leibniz Rule)**.** *Let f and g be n-times differentiable functions. Then fg is also n-times differentiable and*

$$(fg)^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}(x) g^{(k)}(x)$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Proof. Using a proof by induction argument we have the base case that

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x),$$

which is the standard product rule for derivatives.

Now assume that the statement is true for some natural number j. Then

$$\begin{aligned}
(fg)^{(j+1)}(x) &= \frac{d}{dx}\left((fg)^{(j)}(x)\right) \\
&= \frac{d}{dx}\left(\sum_{k=0}^{j}\binom{j}{k} f^{(j-k)}(x)g^{(k)}(x)\right) \\
&= \left(\sum_{k=0}^{j}\binom{j}{k}\left(\frac{d}{dx} f^{(j-k)}(x)g^{(k)}(x)\right)\right) \\
&= \left(\sum_{k=0}^{j}\binom{j}{k}\left(f^{(j-k+1)}(x)g^{(k)}(x) + f^{(j-k)}(x)g^{(k+1)}(x)\right)\right) \\
&= \left(\sum_{k=0}^{j}\binom{j}{k} f^{(j-k+1)}(x)g^{(k)}(x)\right) + \left(\sum_{k=0}^{j}\binom{j}{k} f^{(j-k)}(x)g^{(k+1)}(x)\right).
\end{aligned}$$

We can then re-index the second summation in order to have the order of the derivatives of f and g to align* between the two summations and we have that

$$\begin{aligned}
(fg)^{(j+1)}(x) &= \left(\sum_{k=0}^{j}\binom{j}{k} f^{(j-k+1)}(x)g^{(k)}(x)\right) + \left(\sum_{k=1}^{j+1}\binom{j}{k-1} f^{(j+1-k)}(x)g^{(k)}(x)\right) \\
&= \binom{j}{0} f^{(j+1)}(x)g^{(0)}(x) + \left(\sum_{k=0}^{j}\left(\binom{j}{k} + \binom{j}{k-1}\right) f^{(j+1-k)}(x)g^{(k)}(x)\right) \\
&\quad + \binom{j}{j} f^{(0)}(x)g^{(j+1)}(x)
\end{aligned}$$

Since

$$\binom{j}{0} = \binom{j+1}{0} = 1 \quad \text{and} \quad \binom{j}{j} = \binom{j+1}{j+1} = 1$$

and since

$$\begin{aligned}
\binom{j}{k} + \binom{j}{k-1} &= \frac{j!}{k!(j-k)!} + \frac{j!}{(k-1)!(j-(k-1))!} \\
&= \frac{j!\cdot(j-k+1)}{k!(j-k+1)!} + \frac{j!(k)}{k!(j-k+1)!} \\
&= \frac{j!(j+1)}{k!((j+1)-k)!} = \frac{(j+1)!}{k!((j+1)-k)!} \\
&= \binom{j+1}{k}
\end{aligned}$$

we have that

$$(fg)^{(j+1)}(x) = \sum_{k=0}^{j+1}\binom{j+1}{k} f^{((j+1)-k)}(x)g^{(k)}(x)$$

and so the statement is true for $j+1$.

Therefore, by induction we have that

$$(fg)^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)}(x) g^{(k)}(x)$$

is true for all n. □

Proof of Theorem 8.1. The factorization of p into $p(x) = (x-c)^m \cdot q(x)$ is a direct consequence of Theorem 7.17.

Let $f(x) = (x-c)^m$. Then multiple applications of the chain rule gives us that

$$f^{(j)}(x) = \begin{cases} \frac{m!}{(m-j)!}(x-c)^{m-j} & \text{if } 0 \le j \le m \\ 0 & \text{if } j > m \end{cases}$$

Therefore, using the General Leibniz Rule we have that for $0 \le n \le m$,

$$p^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} \frac{m!}{(m-n+k)!}(x-c)^{m-n+k} q^{(k)}(x)$$

and so

$$p^{(n)}(c) = \begin{cases} 0 & \text{if } 0 \le n < m \\ q(c) & \text{if } n = m \end{cases}$$

giving us the result. □

The key component of this theorem is that we can further understand how the graphs of polynomials behave near their zeros. If we let

$$p(x) = x^6 + x^5 - 11x^4 - 13x^3 + 26x^2 + 20x - 24,$$

we can write this in factored form and see that

$$p(x) = (x-1)^2(x+2)^3(x-3)$$

has a zero of order 1 at $x = 3$, a zero of order 2 at $x = 1$, and a zero of order 3 at $x = -2$. From the theorem we have that near $x = 3$,

$$p(x) \approx (3-1)^2(3+2)^3(x-3) = 500(x-3).$$

Similarly, near $x = 1$,

$$p(x) \approx -54(x-1)^2$$

and near $x = -2$,

$$p(x) \approx -45(x+2)^3.$$

By graphing these polynomials using computer software we can see that these approximations for the polynomial behavior near the zeros is much more accurate than the standard linear, quadratic, or cubic descriptions.

8.6.1.1 Complex Zeros

If $p(x)$ is a polynomial in $\mathbb{R}[x]$, the factors of this polynomial that correspond to complex zeros come in conjugate pairs. This means that if $(x-(a+bi))^j$ is a factor, then so is $\overline{(x-(a+bi))^j} = (x-(a-bi))^j$. While there is no direct correspondence between complex zeros and properties of the graph of a polynomial, If there are local minimum and maximum values for a polynomial for which the graph of the polynomial does not cross the horizontal axis, it implies that there are complex zeros of the polynomial.

8.6.2 Long-Term Behavior

When we study the long-term behavior of polynomials we focus on the standard form of the polynomial

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n \quad \text{with } a_n \neq 0.$$

When we look at the ratio of $p(x)$ and its leading term $a_n x^n$ we have

$$\frac{p(x)}{a_n x^n} = \frac{a_0}{a_n}\frac{1}{x^n} + \frac{a_1}{a_n}\frac{1}{x^{n-1}} + \cdots \frac{a_{n-1}}{a_n}\frac{1}{x} + 1.$$

So we see that

$$\lim_{x\to\pm\infty} \frac{p(x)}{a_n x^n} = 1$$

meaning that these two functions act very similarly as $|x|$ gets large.

This means that the graph of our polynomial

$$p(x) = x^6 + x^5 - 11x^4 - 13x^3 + 26x^2 + 20x - 24$$

will look very similar to the graph of $y = x^6$ when zoomed out to study its long-term behavior.

8.6.3 Polynomial Identities

Just like the difference of squares and perfect square for quadratic polynomials, there are some identities of polynomials that are very useful.

If $p(x) = 1 + x + x^2 + \cdots + x^{n-1}$ is multiplied by $(x-1)$ then one sees that

$$\frac{x^n - 1}{x - 1} = 1 + x + x^2 + \cdots + x^{n-1}.$$

This is the basis for the series

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x},$$

along with understanding the roots of unity for complex numbers.

The identity

$$(x^2+y^2)^2 = x^4 + 2x^2y^2 + y^4 = x^4 - 2x^2y^2 + y^4 + 4x^2y^2 = (x^2-y^2)^2 + (2xy)^2$$

can be used to generate Pythagorean triples. For instance, if $x = 2$ and $y = 1$ the identity becomes $5^2 = 3^2 + 4^2$. If $x = 5$ and $y = 4$ we have the triple $41^2 = 9^2 + 40^2$.

Finally, we have the identity

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}, \quad \text{with } \binom{n}{k} = \frac{n!}{k!(n-k)!},$$

which can be found using Pascal's triangle.

Related Content Standards

- (HSA.APR.4) Prove polynomial identities and use them to describe numerical relationships. *For example, the polynomial identity* $(x^2+y^2)^2 = (x^2-y^2)^2 + (2xy)^2$ *can be used to generate Pythagorean triples.*
- (HSA.APR.5) Know and apply the Binomial Theorem for the expansion of $(x+y)^n$ in powers of x and y for a positive integer n, where x and y are any numbers, with coefficients determined for example by Pascal's Triangle.

8.6.4 Exercises

1. For each of the following polynomials,

   - What are the zeros of f?
   - For what values of x is $f(x) > 0$?
   - For what values of x is $f(x) < 0$?
   - What is $\lim_{x\to\infty} f(x)$?
   - What is $\lim_{x\to-\infty} f(x)$?

   a. $f(x) = (3x+5)(x-1)(7x+3)(x-5)$
   b. $f(x) = (2x-1)(3x+5)(x+1)$
   c. $f(x) = (x-1)(4x+3)(3x-8)$
   d. $f(x) = (2x+1)(3x-6)(x-6)$

2. For which values of k does the equation

$$x^4 - 4x^2 + x + k = 0$$

   have four distinct real solutions?

3. If $f(x) = 3x^2$, what are all real values of a and b for which the graph of $g(x) = ax^2 + b$ is below the graph of $f(x)$ for all values of x?

4. Consider the function $f(x) = x^4 - 25x^2 - 36x$, which has one x-intercept at $(-4, 0)$. Find all the other zeros of the function algebraically.

5. Determine the long-term behavior of the following polynomials.

$$\left(\text{Find } \lim_{x\to\infty} f(x) \quad \text{and} \quad \lim_{x\to-\infty} f(x).\right)$$

   a. $f(x) = 2x^3 - 3x^2 + 4x - 1$
   b. $f(x) = -3x^4 + 2x^2 - 3x + 2$
   c. $f(x) = -5x^5 + 3x^4 + 2x^3 - 10x + 15$
   d. $f(x) = 10x^6 + 7x^3 - 1$

6. For each of the zeros of the following polynomials, find the polynomial approximation of the same degree as the zero near the zero. (If $f(x)$ has a zero of order m at $x = c$, find the Taylor polynomial of degree m centered at $x = c$, $a_c(x-c)^m$.)

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

7. Assume the polynomial f has exactly two local maxima and one local minimum, and that these are the only critical points of f.

   a. Sketch a possible graph of f.
   b. What is the largest number of zeros f could have? (Explain your answer.)
   c. What is the least number of zeros f could have? (Explain your answer.)
   d. What is the least number of inflection points f could have?
   e. What is the smallest degree f could have?
   f. Find two possible formulas for $f(x)$ with two different degrees.

8. Given the function $f(x) = -x^4 - 3x^3 + 101x^2 + 543x + 940$, use a graphing calculator to do the following:

a. Find the zeros of the function. Round any zeros you find to the nearest tenth. Explain how you found your answer step-by-step, as if you were explaining to a student who does not know how to use a graphing calculator.
b. Identify any local minima or maxima of the graph of the function as ordered pairs. Indicate which are local minima and which are local maxima. Round any minima or maxima to the nearest tenth. Explain how you found your answer step-by-step.
c. Find the range of the function. Round the numbers in your answer to the nearest tenth. Explain how you found your answer step-by-step.
d. Sketch the graph of the function. List a viewing window that shows all the information you found in the other parts of the problem. Explain why you chose your viewing window.

9. For each of the following graphs, give a possible algebraic representation for $p(x)$.

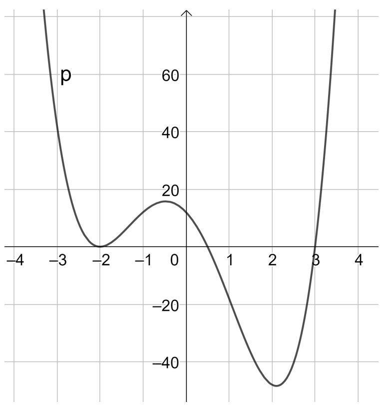


Figure 8.16: Graph 1

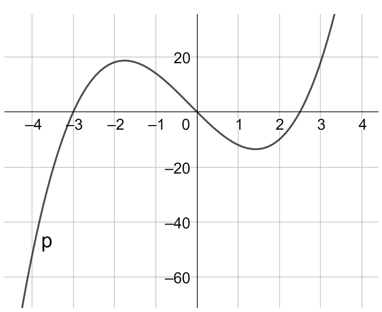


Figure 8.17: Graph 2

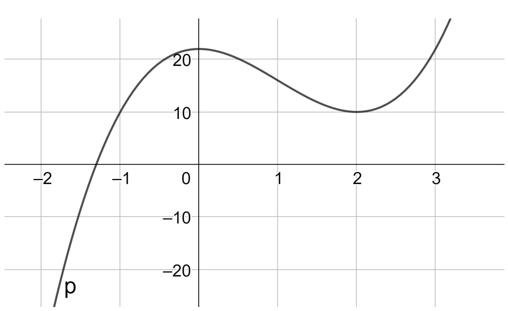


Figure 8.18: Graph 3

8.7 Rational Functions

A rational function is of the form

$$f(x) = \frac{p(x)}{q(x)}$$

for some polynomials p and q with real-valued coefficients. If the degree of q is zero, then the rational function is just a polynomial. So we will assume throughout this section that the degree of q is at least one. In the same way that the different representations of a polynomial help us to understand different aspects of the polynomial, different representations of the rational function will help us to better understand properties of the rational function.

8.7.1 Short-Term Behavior

In order to understand the short-term behavior of a rational function, it is helpful to write the function in its factored form

$$f(x) = \frac{p(x)}{q(x)} = A\frac{(x-a_1)^{j_1}(x-a_2)^{j_2}\cdot(x-a_n)^{j_n}}{(x-b_1)^{k_1}(x-b_2)^{k_2}\cdots(x-b_m)^{k_m}}$$

where the $a_i \in \mathbb{C}$ are the zeros of degree j_i of the numerator (and thus the zeros of the function), the $b_i \in \mathbb{C}$ are zeros of the denominator, and A is a real number that allows the numerator and denominator to both be monic polynomials. Since the polynomials have real-valued coefficients, any factors corresponding to non-real zeros come in conjugate pairs. The b_i that are real-valued correspond with values that have been removed from the domain of the function. Similarly, the a_i that are real-valued correspond to the horizontal intercepts for the rational function.

8.7.1.1 Common Factors

When the numerator and denominator share a factor, $(x - a_0)$,we need to understand how this affects the rational function.

If the degree of the shared factor is larger in the numerator than in the denominator, the rational function is undefined at the zero of the factor, but has a limit of zero near this value. We can also see that the rational function is equivalent to the rational function where $(x - a_0)$ is only a factor in the numerator whose degree is the difference between the degrees of the factor in the numerator and denominator of the original rational function, with the zero of the factor removed from the domain of the new function. For example,

$$f(x) = \frac{(x-1)^3}{(x-1)^1} \text{ is equivalent to } g(x) = (x-1)^2, \text{ for } x \in \mathbb{R} \setminus \{1\}.$$

If the degree of the shared factor is larger in the denominator than in the numerator, the rational function is equivalent to the rational function where $(x - a_0)$ is only a factor in the denominator whose degree is the difference between the degrees of the factor in the denominator and numerator of the original rational function. For example,

$$f(x) = \frac{(x-1)^1}{(x-1)^3} \text{ is equivalent to } g(x) = \frac{1}{(x-1)^2}.$$

If the shared factor has the same degree in the numerator and denominator, then this is a removable singularity and the factor can be removed from both the numerator and denominator with the removal of the zero of the factor from the new rational function.

For the remainder of the section we will assume that the numerator and denominator do not have a common factor.

8.7.1.2 Behavior near Zeros

The factors of the numerator of the rational function determine the zeros of the polynomial. If $(x - a_0)$ is a factor of the numerator of degree j_0, and if $a_0 \in \mathbb{R}$, then a_0 is a zero of the rational function (thus a_0 is a horizontal intercept of the graph of the rational function) and we can rewrite the rational function as

$$f(x) = (x - a_0)^{j_0} g(x)$$

where g is a rational function without $(x - a_0)$ as a factor of either the numerator or denominator (See Theorem 8.1). Using properties from analysis (similarly to Theorem 8.1) we can determine that near $x = a_0$ that

$$f(x) \approx g(a_0)(x - a_0)^{j_0}.$$

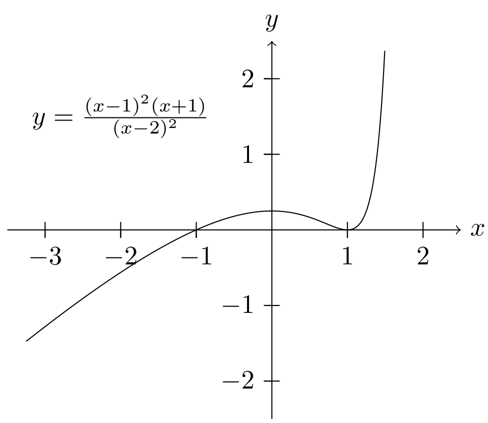
Example 8.1. To better understand this concept let us consider

$$f(x) = \frac{(x-1)^2(x+1)}{(x-2)^2}.$$

We see that the function has horizontal intercepts at -1 and 1. We can also see that near $x = -1$, the function is very similar to

$$g(x) = (x+1) \cdot \frac{((-1)-1)^2}{((-1)-2)^2} = \frac{4}{9}(x+1).$$

Similarly, near $x = 1$ the function behaves very similarly to $h(x) = 2(x-1)^2$.



8.7.1.3 Behavior near Singularities

In order to determine the domain of the rational functions we consider the factors of the denominator of the function. If $(x - b_0)$ is a factor in the denominator of the rational function, and $b_0 \in \mathbb{R}$, then b_0 is not in the domain of the function. If the degree of the factor is k_0, then we see that the function can be rewritten as

$$f(x) = \frac{1}{(x - b_0)^{k_0}} g(x)$$

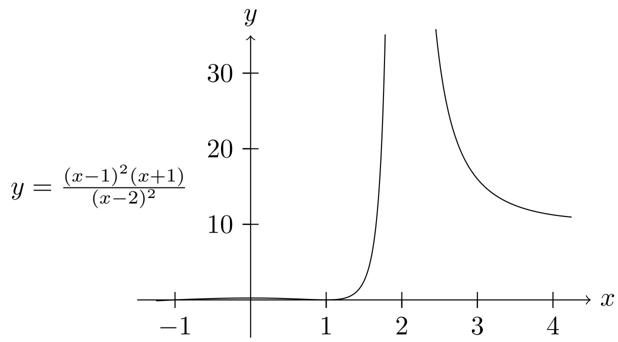
for a rational function g with b_0 in the domain of g and $g(b_0) \neq 0$. So near $x = b_0$,

$$f(x) \approx g(b_0) \frac{1}{(x - b_0)^{k_0}}.$$

Example 8.2. Returning to our example of $f(x) = \frac{(x-1)^2(x+1)}{(x-2)^2}$, we can see that near $x = 2$ that

$$f(x) \approx \frac{3}{(x-2)^2},$$

as evidenced in the following graph.



This understanding of short-term behavior can help us to simplify certain models near given points, choose appropriate window sizes to have technology graph a function, and be able to more easily determine if the graph of a function generated by a program is accurate.

8.7.2 Long-Term Behavior

In order to understand the long-term behavior of rational functions we start by looking at rational functions where the degree of the numerator is less than the degree of the denominator. Such functions can be generalized as

$$f(x) = \frac{a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n}{b_0 + b_1 x + \cdots + b_n x^n + \cdots + b_{n+r} x^{n+r}}.$$

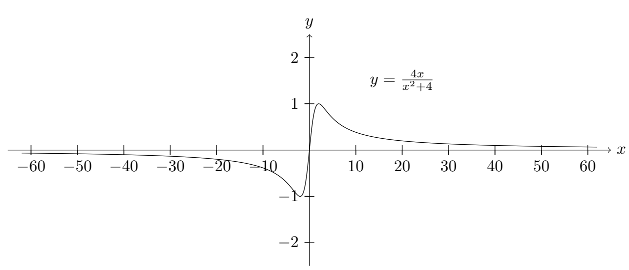
Then by multiplying the numerator and denominator by $x^{-(n+r)}$ we have

$$f(x) = \frac{a_0 \frac{1}{x^{n+r}} + a_1 \frac{1}{x^{n+r-1}} + \cdots + a_n \frac{1}{x^r}}{b_0 \frac{1}{x^{n+r}} + b_1 \frac{1}{x^{n+r-1}} + \cdots + b_n \frac{1}{x} + \cdots + b_{n+r}}$$

and we can see that as $|x|$ gets large that most of the terms converge to 0. Thus,

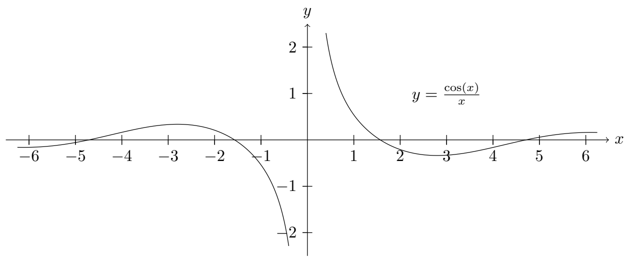
$$\lim_{x \to \pm\infty} f(x) = \frac{0}{b_{n+r}} = 0.$$

So we see that if the degree of the numerator is less than the degree of the denominator that the graph of the function approaches the horizontal axis as $|x|$ gets large. This is often called a horizontal asymptote of the function since the function asymptotically approaches the horizontal axis. A challenge that many students have with this terminology is that they have been told that a function cannot cross an asymptote. While that is true for vertical asymptotes, it is not true for horizontal asymptotes as seen in the graph of the function below.



In fact, a function can cross its “horizontal asymptote” an infinite number of times. For example,

$$f(x) = \frac{\cos(x)}{x}.$$



We can then build on this understanding to study the long-term behavior of rational functions with the degree of the numerator greater than or equal to the degree of the denominator. So if $f(x)$ is a rational function with numerator $a(x)$ and denominator $b(x)$ we have from the Division Algorithm for Polynomials 7.11 that there exist unique polynomials $q(x)$ and $r(x)$ such that

$$a(x) = b(x)q(x) + r(x)$$

with the degree of $r(x)$ less than the degree of $b(x)$. Dividing both sides of this equation by $b(x)$ we have

$$f(x) = \frac{a(x)}{b(x)} = q(x) + \frac{r(x)}{b(x)}.$$

From the conversation above we can see that

$$\lim_{x \to \pm\infty} \frac{r(x)}{b(x)} = 0$$

and that $f(x)$ asymptotically approaches $q(x)$ as $|x| \to \infty$. So if the degree of the numerator is one plus the degree of the denominator then the quotient would be a linear function. Such asymptotes are often called "skew asymptotes.' '

Example 8.3. Returning to our example of $f(x) = \frac{(x-1)^2(x+1)}{(x-2)^2}$, we can rewrite the function using the division algorithm as

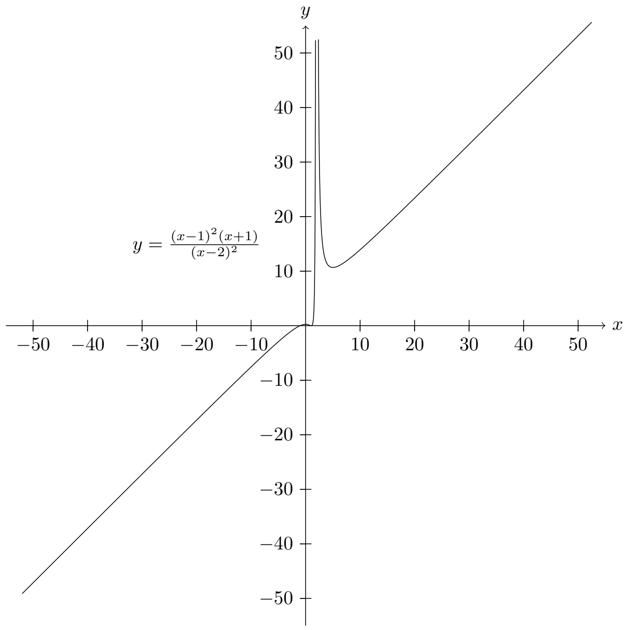
$$f(x) = x + 3 + \frac{7x - 11}{(x-2)^2}.$$

8.7.3 Partial Fraction Decomposition

Another way to represent rational functions whose numerator has a lower degree than the denominator is a **partial fraction decomposition**. This decomposition allows the rational function to be written as the sum of rational functions whose denominators are irreducible and whose numerators are of lower degree than the denominator. For instance,

$$\frac{3x-1}{x^2-2x-3} = \frac{1}{x+1} + \frac{2}{x-3} \quad \text{or} \quad \frac{3x^2-x+1}{x^3+x} = \frac{2x-1}{x^2+1} + \frac{1}{x}.$$

There are limited uses of the partial fraction decomposition in the K-12 curriculum. Hence, it was not included in the Common Core standards. However, this representation of a rational function is essential in the area of integration of rational functions. It is also useful to find Taylor (and Laurent) series expansions, and to solve problems in differential equations involving the Laplace transform. Because of this, we recommend this content first being taught during the section of integration of rational functions in Calculus. In the high school curriculum, this is during the BC portion of an Advanced Placement Calculus AB/BC course.



8.7.4 Exercises

1. Consider the polynomials $a(x) = 2x^3 + 3x^2 - 5x - 7$ and $b(x) = x^2 + 1$.

   a. Perform polynomial division of $a(x)$ by $b(x)$. Call the quotient polynomial $q(x)$ and the remainder polynomial $r(x)$.
   b. Use the quotient and remainder to describe the behavior of $f(x)$ when $|x|$ is large.

2. Let $f(x) = \dfrac{x^3 - 1}{x^3 + 1}$.

   a. What is the domain of f?
   b. Determine the singular points of f (that is, the points at which f is undefined) and whether f has any vertical asymptotes at these points.
   c. What is the range of f?
   d. At what points does the graph of f intersect the horizontal-axis?
   e. Find any global or local maxima or minima.
   f. Determine the intervals on which f is increasing and those intervals on which f is decreasing.
   g. Describe what happens to $f(x)$ as $x \to \infty$ and as $x \to -\infty$.
   h. What is a power series of f?

3. List and describe the behavior near each of the zeros and singularities for the following rational functions.

   a. $f(x) = \dfrac{x+3}{x^2 + 2x - 3}$
   b. $f(x) = \dfrac{x^2 - 2x - 24}{x^2 + 10x + 24}$
   c. $f(x) = \dfrac{25x^2 - 80x + 64}{192 - 75x^2}$
   d. $f(x) = \dfrac{12x^7 - 3x^3}{4x^4 - 16x}$
   e. $f(x) = \dfrac{6x - 12x^9}{3x^3 + 7}$

f. $f(x) = \dfrac{x^2 + 15x + 56}{x + 5}$

g. $f(x) = \dfrac{x^2 + 2x - 15}{x^2 + 10x + 25}$

h. $f(x) = \dfrac{x^2 + 8x + 16}{x^2 + 6x + 8}$

i. $f(x) = \dfrac{x^5 + 3x^3 + 2x^2 + 4x + 1}{x^4 + 5}$

4. Describe the long-term behavior for the following rational functions.

a. $f(x) = \dfrac{x + 3}{x^2 + 2x - 3}$

b. $f(x) = \dfrac{x^2 - 2x - 24}{x^2 + 10x + 24}$

c. $f(x) = \dfrac{25x^2 - 80x + 64}{192 - 75x^2}$

d. $f(x) = \dfrac{12x^7 - 3x^3}{4x^4 - 16x}$

e. $f(x) = \dfrac{6x - 12x^9}{3x^3 + 7}$

f. $f(x) = \dfrac{x^2 + 15x + 56}{x + 5}$

g. $f(x) = \dfrac{x^2 + 2x - 15}{x^2 + 10x + 25}$

h. $f(x) = \dfrac{x^2 + 8x + 16}{x^2 + 6x + 8}$

i. $f(x) = \dfrac{x^5 + 3x^3 + 2x^2 + 4x + 1}{x^4 + 5}$

5. Find an algebraic representation for a function, f, whose graph has all of the following attributes.

- $f(2) = 0$
- $f(\frac{1}{2}) = 0$
- $f(-3) = 0$
- $\lim\limits_{x \to \frac{2}{3}^+} f(x) = \infty$
- $\lim\limits_{x \to 1^-} f(x) = \infty$
- As $x \to \pm\infty$, $f(x) \approx \dfrac{10}{3}x^2 + \dfrac{77}{9}x - \dfrac{242}{27}$

6. Write the rational functions below in their partial fraction decomposition over the real numbers.

a. $f(x) = \dfrac{x^3 + x^2 + 1}{x^2 - x}$

b. $f(x) = \dfrac{x^4}{x^4 - 1}$

c. $f(x) = \dfrac{4x^2 - 3x + 2}{4x^2 - 4x + 3}$

d. $f(x) = \dfrac{x^4 - 2x^2 + 4x + 1}{x^3 - x^2 - x + 1}$

e. $f(x) = \dfrac{2x^2 - x + 4}{x^3 + 4x}$

8.8 Trigonometric Functions

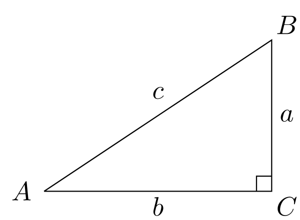
The area of trigonometry comes from the combination of τριγωνς (triangle) and μετρον (measure) and is the study of the measurements of triangles. The study of the measurements of sides of triangles was known to the ancient Egyptians and Babylonians for the purpose of constructions and land surveys [Boyer and Merzbach, 1991, p. 158]. The Greeks expanded on this work with many of our standard trigonometric properties included in Euclid's Elements. With the growth of astronomy and navigation in the 1500's trigonometry was expanded to include more identities. The current uses of trigonometric functions include optics, acoustics, electrical engineering, image compression, and computer-generated imagery (CGI).

8.8.1 Definitions

Beginning in fourth grade, students are introduced to right triangles, with a cyclical study of these triangles continuing through middle school, and into high school.

Related Content Standards

- (4.G.2) Classify two-dimensional figures based on the presence or absence of parallel or perpendicular lines, or the presence or absence of angles of a specified size. Recognize right triangles as a category, and identify right triangles.
- (6.G.1) Find the area of right triangles, other triangles, special quadrilaterals, and polygons by composing into rectangles or decomposing into triangles and other shapes; apply these techniques in the context of solving real-world and mathematical problems.
- (8.G.5) Use informal arguments to establish facts about the angle sum and exterior angle of triangles, about the angles created when parallel lines are cut by a transversal, and the angle-angle criterion for similarity of triangles.
- (HSG.SRT.6) Understand that by similarity, side ratios in right triangles are properties of the angles in the triangle, leading to definitions of trigonometric ratios for acute angles.



Our first definitions for the trigonometric functions are based on the properties of right triangles. As such, the initial domain of the trigonometric functions are defined in degrees on $(0°, 90°)$. For an angle A, we define the sine of angle A to be the ratio of the measure of the opposite side, a, and the hypotenuse, c,

$$\sin(A) = \frac{a}{c}.$$

The cosine of angle A is the ratio of the measures of the adjacent side, b, and the hypotenuse, c, and the tangent of the angle A is the ratio of the measures of the opposite side, a, and the adjacent side, b,

$$\cos(A) = \frac{b}{c} \quad \text{and} \quad \tan(A) = \frac{a}{b}.$$

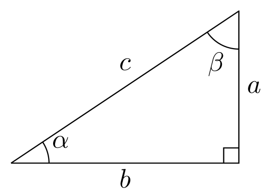
Similarly for angle B,

$$\sin(B) = \frac{b}{c}, \quad \cos(B) = \frac{a}{c}, \quad \text{and} \quad \tan(B) = \frac{b}{a}.$$

We also define the additional trigonometric functions as follows:

$$\text{secant of } A = \sec(A) = \frac{1}{\cos(A)} = \frac{c}{b}$$
$$\text{cosecent of } A = \csc(A) = \frac{1}{\sin(A)} = \frac{c}{a}$$
$$\text{cotangent of } A = \cot(A) = \frac{1}{\tan(A)} = \frac{b}{a}$$

While these definitions are sufficient, it is less ambiguous to define the angle, not by the vertex, but by the amount between two line segments.



$$\alpha + \beta = 90^\circ \qquad \alpha = 90^\circ - \beta \qquad \beta = 90^\circ - \alpha$$
$$\cos(\alpha) = \sin(\beta) = \frac{b}{c} \qquad \tan(\alpha) = \cot(\beta) = \frac{a}{b} \qquad \sec(\alpha) = \csc(\beta) = \frac{c}{b}$$
$$\sin(\alpha) = \cos(\beta) = \frac{a}{c} \qquad \cot(\alpha) = \tan(\beta) = \frac{b}{a} \qquad \csc(\alpha) = \sec(\beta) = \frac{c}{a}$$

8.8.1.1 Radians and Degrees

The use of 360° to measure angles is likely related to the number of days (365) it takes for the Earth to rotate around the sun. With the advent of calculus it became apparent that a different technique of measuring an angle based on the arc length subtended by the angle was needed. Shortly after his death in 1716 at the age of 33, Roger Cotes published the first definition of the radian angle measurement [1722]. (It was in this same work that he wrote $ix = ln(\cos(x) + i\sin(x))$, the precursor to Euler's formula.)

A radian is defined to be the angle subtended from by a circular arc whose length is the same as the radius. This means that there are 2π radians in a circle, a right angle is $\frac{\pi}{2}$ radians, and a straight line is π radians.

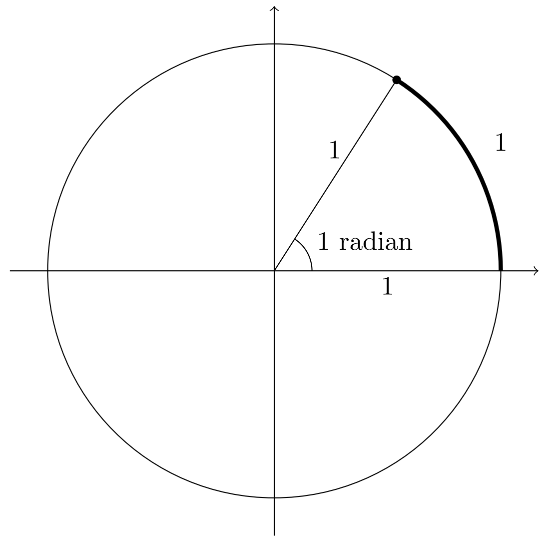
In order to convert an angle from degrees to radians, one would multiply by $\frac{\pi}{180^\circ}$. And to convert from radians to degrees would be necessitate multiplying by the reciprocal.

One of the primary benefits of using radians instead of degrees is that for angles near 0, $\sin(\alpha) \approx \alpha$. This simplifies the derivatives of the trigonometric functions and makes for a cleaner series expansion for the trigonometric functions.

In addition, to find the arclength of an angle in radians, one only needs to multiply by the radius, ($l = \alpha r$). To find the arclength in for an angle given in degrees we have $l = \alpha\frac{\pi r}{180^\circ}$.

Related Content Standards

- (HSF.TF.1) Understand radian measure of an angle as the length of the arc on the unit circle subtended by the angle.



8.8.1.2 Unit Circle

Combining the definitions of the trigonometric functions for right triangles with the angles defined by radians we can extend the domain of trigonometric functions to all possible angles using the unit circle. For angles $\alpha \in \left(0, \frac{\pi}{2}\right)$, we plot a point on the unit circle that is swept out by that angle from the positive horizontal axis. We then form a right triangle from that point by drawing a line segment from the point that is perpendicular to the horizontal axis and draw a line segment from the point to the origin. We can then see that the coordinates of the point are $(\cos(\alpha), \sin(\alpha))$. This also means that the distance along the unit circle from $(1, 0)$ to this point is α.

Then, for any angle α we define $\sin(\alpha)$ and $\cos(\alpha)$ to be given by the coordinates on the unit circle whose distance from the point $(1, 0)$ along the circle in a counter-clockwise direction is α.

Related Content Standards

- (HSF.TF.2) Explain how the unit circle in the coordinate plane enables the extension of trigonometric functions to all real numbers, interpreted as radian measures of angles traversed counterclockwise around the unit circle.

With this extended definition of sine and cosine, we can define all of our trigonometric functions on a much larger domain. We can see from these definitions that

$$\sin : \mathbb{R} \to [-1, 1], \ \cos : \mathbb{R} \to [-1, 1], \text{ and}$$

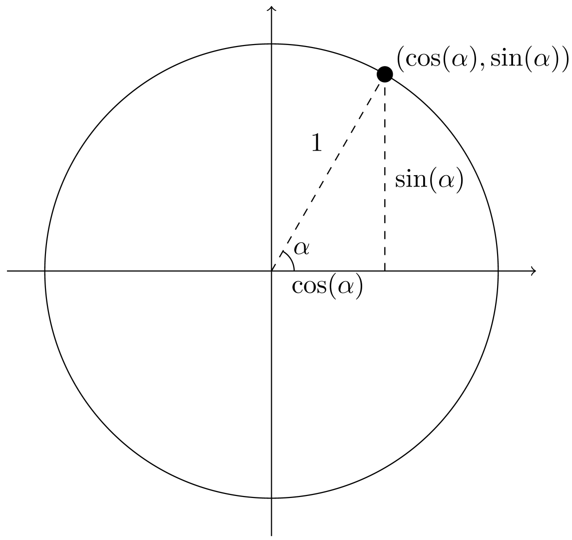
$$\tan : \left(\mathbb{R} \setminus \left\{\frac{\pi}{2} + k\pi : k \in \mathbb{Z}\right\}\right) \to \mathbb{R}.$$

Since $(\cos(\alpha), \sin(\alpha))$ is defined as a point on the unit circle, we know from the Pythagorean theorem that $(\cos(\alpha))^2 + (\sin(\alpha))^2 = 1$. In order to reduce the notation we often write $\cos^2(\alpha)$ and $\sin^2(\alpha)$ to represent the square of the function. With this notation we have

$$\cos^2(\alpha) + \sin^2(\alpha) = 1.$$

This equation can then be rearranged to create the equivalent equations of

$$\sec^2(\alpha) - \tan^2(\alpha) = 1 \quad \text{and} \quad \csc^2(\alpha) - \cot^2(\alpha) = 1.$$



Related Content Standards

- (HSF.TF.8) Prove the Pythagorean identity $\sin^2(\theta) + \cos^2(\theta) = 1$ and use it to find $\sin(\theta)$, $\cos(\theta)$, or $\tan(\theta)$ given $\sin(\theta)$, $\cos(\theta)$, or $\tan(\theta)$ and the quadrant of the angle.

Using the properties of the unit circle we find that the values of the cosine and sine functions repeat with every cycle around the unit circle and so we find that these are periodic functions with a period of 2π,

$$\cos(\alpha) = \cos(\alpha + 2\pi) \quad \text{and} \quad \sin(\alpha) = \sin(\alpha + 2\pi).$$

Since the negative angles are the rotation in the clockwise direction instead of the counter-clockwise direction we have some additional symmetry,

$$\sin(-\alpha) = -\sin(\alpha) \quad \text{and} \quad \cos(-\alpha) = \cos(\alpha),$$

and so we see that the sine function is an odd function and the cosine function is an even function.

Related Content Standards

- (HSF.TF.4) Use the unit circle to explain symmetry (odd and even) and periodicity of trigonometric functions.

8.8.2 Additional Properties

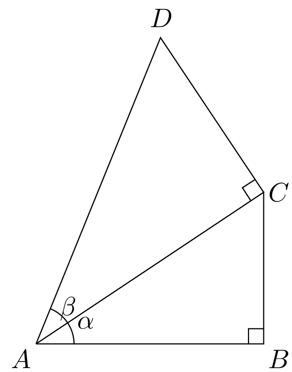
8.8.2.1 Angle Addition Formulas

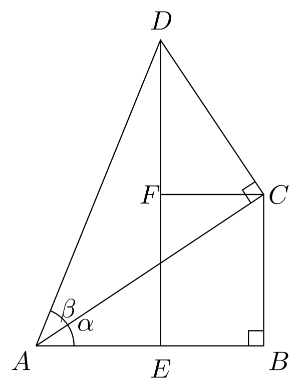
In order to prove the angle addition formula for the sine function we begin by choosing two acute (less than right) angles, α and β, and draw a pair of right triangles so that the hypotenuse of the triangle ABC is a leg of the triangle ACD, as in the figure below.

We can then draw a line segment through D perpendicular to AB, and a another line segment perpendicular to this new line and through the point C.

We now have that

$$\sin(\alpha + \beta) = \frac{DE}{AD}$$





and that $DE = DF + EF$, and since $EF = BC$ we have that

$$\sin(\alpha+\beta) = \frac{DF + BC}{AD} = \frac{DF}{AD} + \frac{BC}{AD}$$

We can also view AC as a transversal to the parallel lines AB and CF so that the angle $\angle ACF = \alpha$ and also $\angle FDC = \alpha$. This means that $\cos(\alpha) = \frac{DF}{CD}$. We can then rewrite the angle addition equation as

$$\sin(\alpha+\beta) = \frac{DF}{CD}\frac{CD}{AD} + \frac{BC}{AC}\frac{AC}{AD} = \cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta).$$

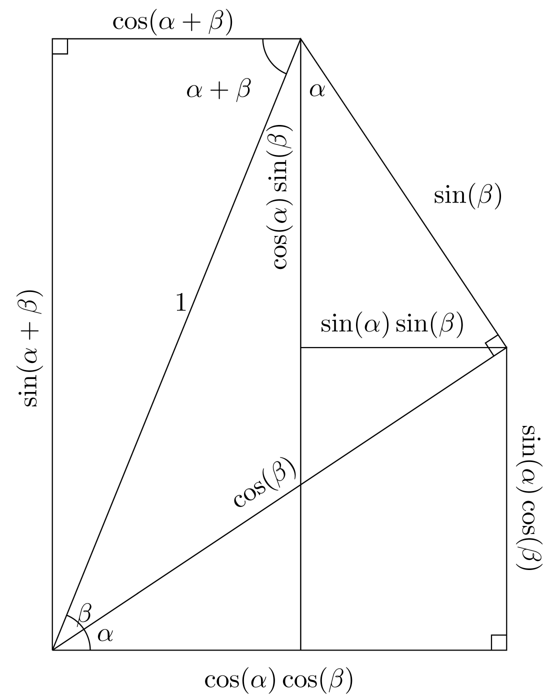
In order to prove the angle addition formula for the cosine function we can see that

$$\begin{aligned}
\cos(\alpha+\beta) &= \sin\left(\frac{\pi}{2} - \alpha - \beta\right) \\
&= \sin\left(\left(\frac{\pi}{2} - \alpha\right) + (-\beta)\right) \\
&= \cos\left(\frac{\pi}{2} - \alpha\right)\sin(-\beta) + \sin\left(\frac{\pi}{2} - \alpha\right)\cos(-\beta) \\
&= \sin(\alpha)\left(-\sin(\beta)\right) + \cos(\alpha)\cos(\beta) \\
&= \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)
\end{aligned}$$

Another proof of both of these identities involves can be constructed from the following diagram.

If the angles are obtuse, then one can use the symmetry of the sine and cosine functions to change the angle addition into a sum of angles that are acute.

Related Content Standards



- (HSF.TF.9) Prove the addition and subtraction formulas for sine, cosine, and tangent and use them to solve problems.

If we use Euler's equation, $e^{iy} = \cos(y) + i\sin(y)$, we can prove the angle addition formulas using properties of exponents and the addition and multiplication operations. We can write $e^{i(\alpha+\beta)}$ in two ways:

$$\begin{aligned} e^{i(\alpha+\beta)} &= e^{i\alpha} \cdot e^{i\beta} = (\cos(\alpha) + i\sin(\alpha))(\cos(\beta) + i\sin(\beta)) \\ &= (\cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)) + i\,(\sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)) \\ e^{i(\alpha+\beta)} &= \cos(\alpha+\beta) + i\sin(\alpha+\beta) \end{aligned}$$

Then by equating the real part and imaginary parts of these two representations we have the angle addition formulas.

Related Content Standards

- (HSG.SRT.10) Prove the Laws of Sines and Cosines and use them to solve problems.
- (HSG.SRT.11) Understand and apply the Law of Sines and the Law of Cosines to find unknown measurements in right and non-right triangles (e.g., surveying problems, resultant forces).

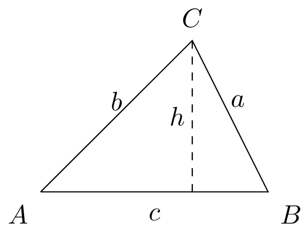
Theorem 8.2 (Law of Sines)**.** *If a triangle has sides of lengths a, b, and c opposite angles A, B, and C, respectively, then*

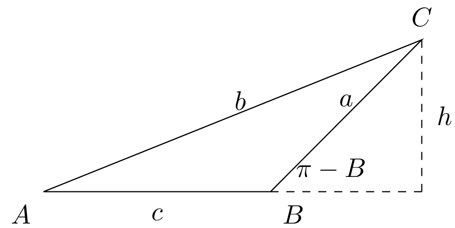
$$\frac{\sin(A)}{a} = \frac{\sin(B)}{b} = \frac{\sin(C)}{c}$$

Proof. If the triangle has a right angle, then the law of sines follows directly from the definition of the sine function. So we may assume that we have a triangle, $\triangle ABC$, without any right angles.

Then $\angle ABC$ can be acute or obtuse.

In each case, draw the altitude from the vertex at C to the side AB. In the acute triangle the altitude lies inside the triangle, while in the obtuse triangle the altitude lies outside the triangle.





Let h be the height of the altitude. For each triangle we see that

$$\frac{h}{b} = \sin(A) \quad \text{and} \quad \frac{h}{a} = \sin(B),$$

since $\frac{h}{a} = \sin(\pi - B) = \sin(B)$ because of the symmetry of the sine function. Thus, solving for h and substituting that into the previous equation gives

$$a\sin(B) = b\sin(A) \quad \text{or equivalently} \quad \frac{\sin(A)}{a} = \frac{\sin(B)}{b}.$$

By a similar argument, drawing the altitude from A to BC gives

$$\frac{\sin(C)}{c} = \frac{\sin(B)}{b}.$$

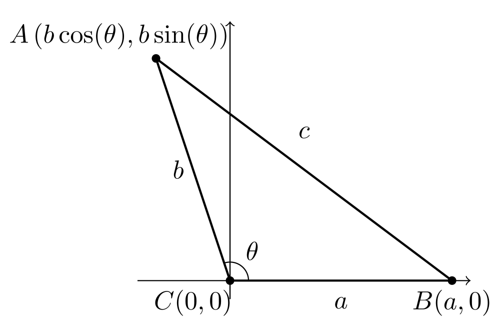
We then combine these to complete our proof. □

Theorem 8.3 (Law of Cosines)**.** *If a triangle has sides of lengths a, b, and c, opposite angles A, B, and C, respectively, then*

$$c^2 = a^2 + b^2 - 2ab\cos(C)$$

There are many different proofs of the Law of Cosines. We include one from the perspective of analytic geometry.

Proof. We are able to create a system of coordinates on the plane in order to align the point C with the origin and point B on the positive real axis at the point $(a, 0$.



We can then use the distance formula to find the distance from A to C as

$$c^2 = (b\cos(\theta) - a)^2 + (b\sin(\theta))^2 = b^2\cos^2(\theta) - 2ab\cos(\theta) + a^2 + b^2\sin^2(\theta)$$

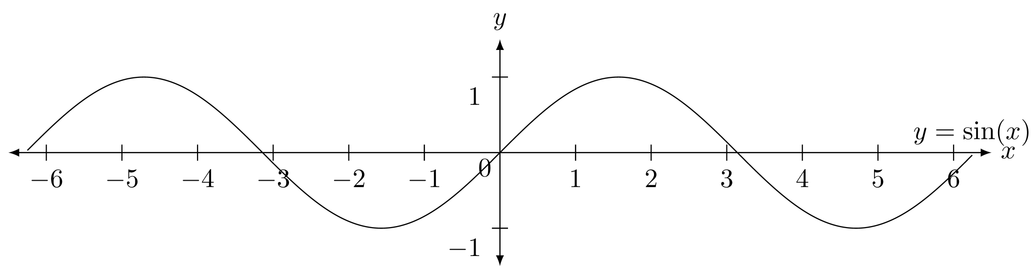
We can then rearrange the terms and use the property that $b^2\cos^2(\theta) + b^2\sin^2(\theta) = b^2$ to see that

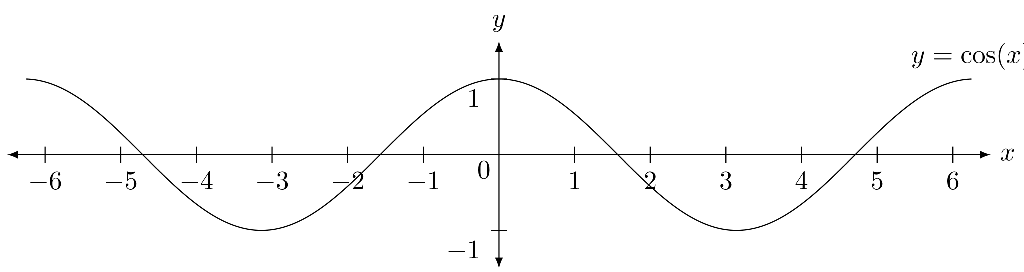
$$c^2 = a^2 + b^2 - 2ab\cos(C)$$

□

8.8.3 Graphs of Trigonometric Functions

Using the definitions of the sine and cosine functions from the unit circle we can sketch their graphs as a function of the angle in radians.





We can then look at the family of functions generated from these two functions. One thing to notice is that the sine function is a horizontal shift of the cosine function, $\sin(x) = \cos\left(x - \frac{\pi}{2}\right)$. Because of this, we will only look at the family of functions generated from the cosine function. This family of functions is all of the functions of the form

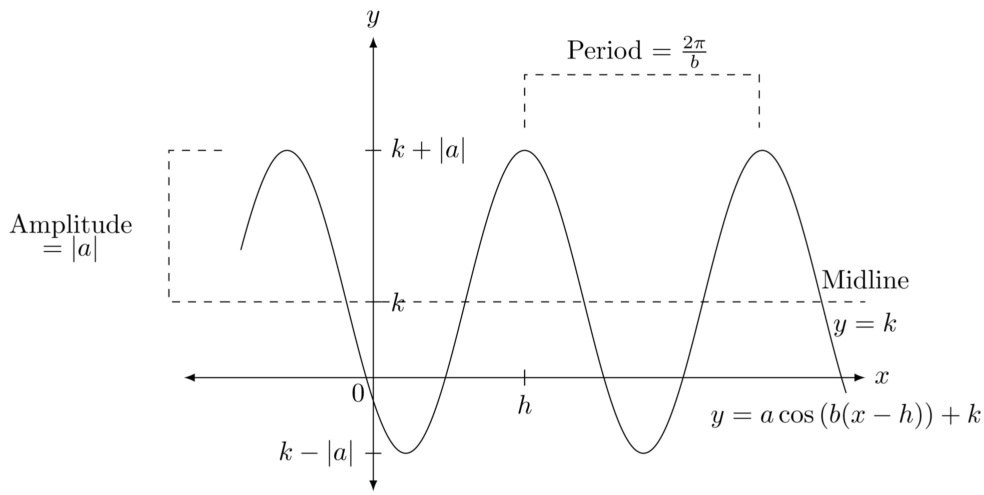
$$f(x) = a\cos\left(b(x - h)\right) + k$$

for real numbers a, b, h, and k.

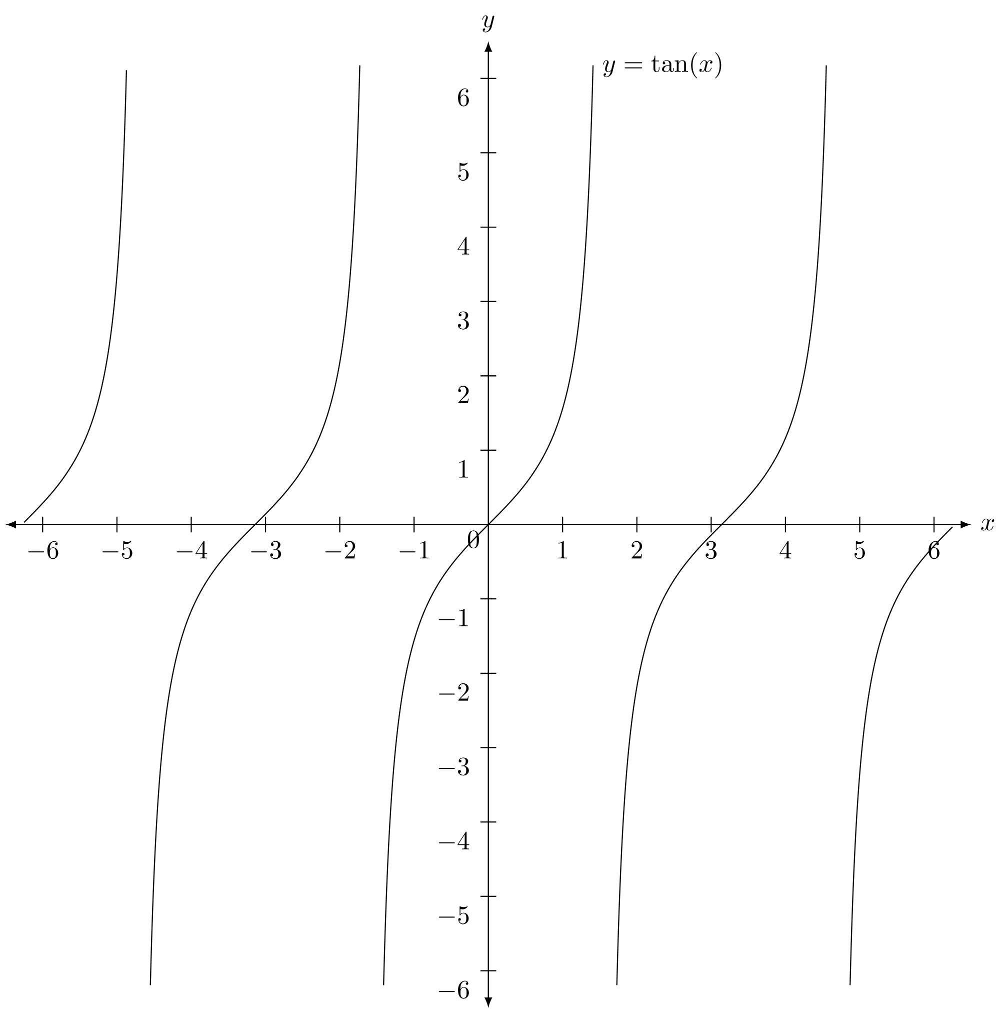
Related Content Standards

- (HSF.TF.5) Choose trigonometric functions to model periodic phenomena with specified amplitude, frequency, and midline.

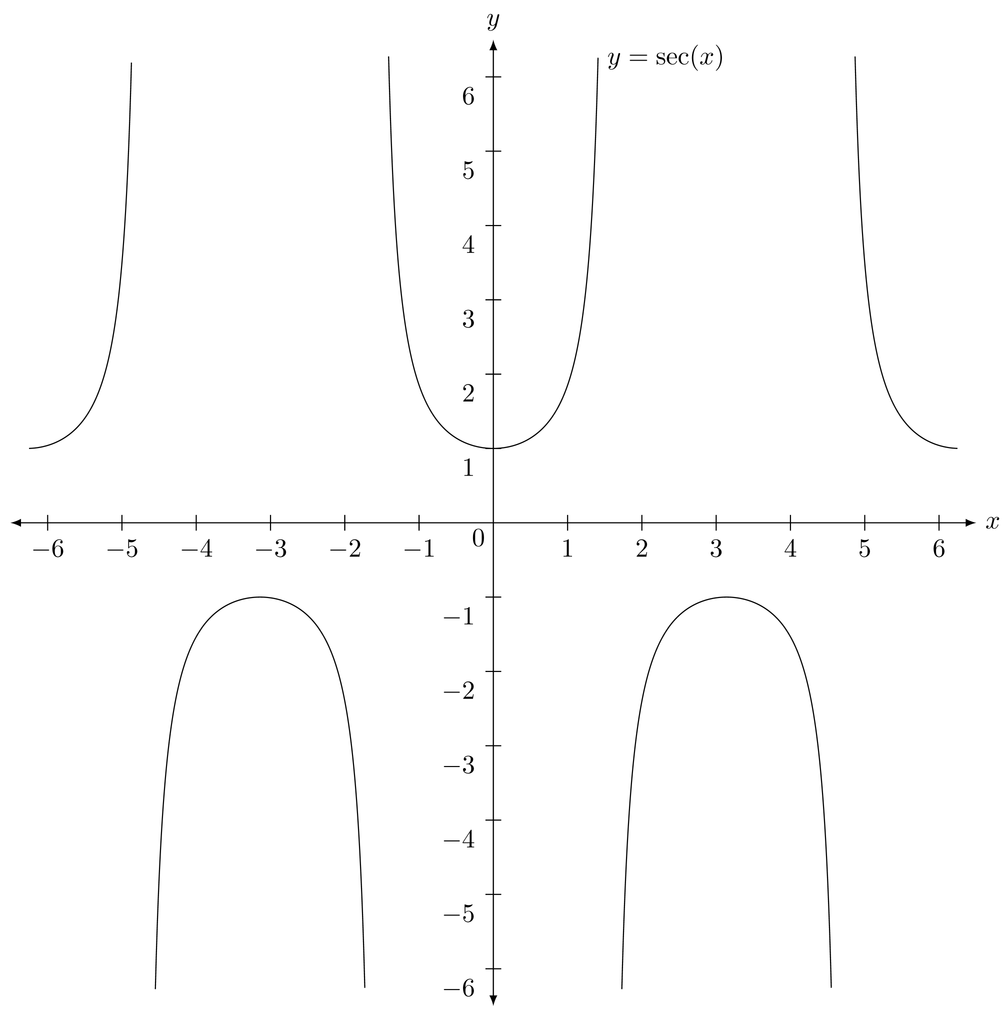
We can also sketch the base graphs of the other trigonometric functions noticing that the domain of these functions will not include certain values as we cannot divide by zero and have a real number. We will give the graphs of the tangent and secant functions, with the graphs of the cotangent and cosecant functions being similar.



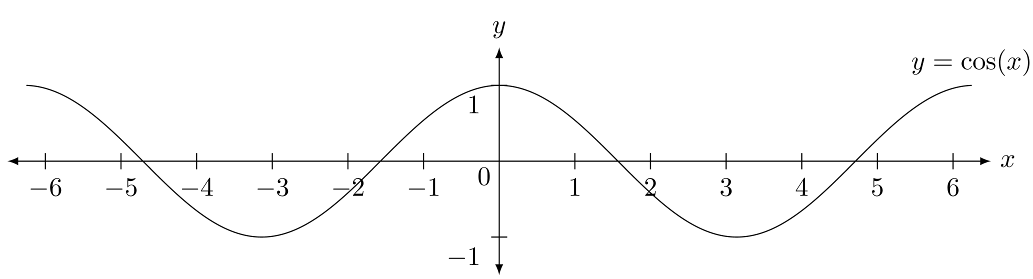
y
Period $= \frac{2\pi}{b}$
$k + |a|$
Amplitude
$= |a|$
Midline
k
$y = k$
x
0
h
$y = a\cos(b(x-h)) + k$
$k - |a|$



y
$y = \tan(x)$
6
5
4
3
2
1
0
x
-6 -5 -4 -3 -2 -1
1 2 3 4 5 6
-1
-2
-3
-4
-5
-6



y
y = sec(x)
6
5
4
3
2
1
0
x
−6 −5 −4 −3 −2 −1 1 2 3 4 5 6
−1
−2
−3
−4
−5
−6



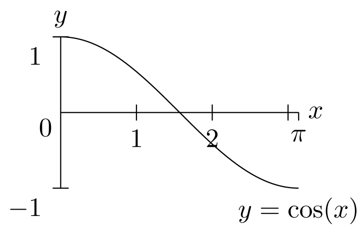
y
y = cos(x)
1
0
x
−6 −5 −4 −3 −2 −1 1 2 3 4 5 6
−1

8.8.4 Inverse Trigonometric Functions

From Section 5.3, we see that a function is invertible if, and only if, it is a bijection. We can look at the graphs of the sine and cosine functions and see that they are definitely not bijections when thought of as functions from $\mathbb{R}$ to $\mathbb{R}$.

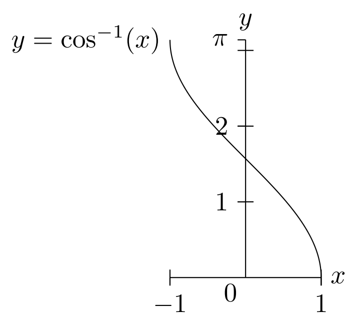
However, if we restrict the domain and co-domain of the functions, these new functions are invertible.

For the cosine function, we restrict the domain to $[0, \pi]$ in order to find the maximum region on which the function is an injection and is onto the range of the original function, $[-1, 1]$. While there are many options for this choice, we choose the option that includes the first quadrant to correspond with the definitions of the functions for right triangles.

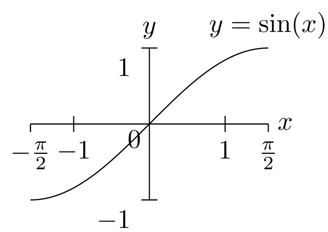


This function is now invertible and so we call its inverse function arcosine and denote it

$$\arccos : [-1, 1] \to [0, \pi] \quad \text{or} \quad \cos^{-1} : [-1, 1] \to [0, \pi].$$



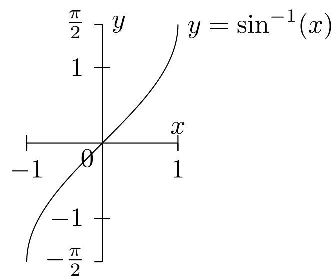
Similarly we will restrict the sine function to the largest domain that includes the first quadrant for which the function is an injection.



We then call the inverse function of this restricted domain function the arcsine and denote it

$$\arcsin : [-1, 1] \to \left[\frac{-\pi}{2}, \frac{\pi}{2}\right] \quad \text{or} \quad \sin^{-1} : [-1, 1] \to \left[\frac{-\pi}{2}, \frac{\pi}{2}\right].$$

We can then say that the arcsine of a number, x, between -1 and 1 is the angle, θ, between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$ so that $\sin(\theta) = x$. With a similar interpretation for the arccosine.



We can similarly restrict the domain and co-domain of the tangent, cotangent, secant, and cosecant functions to be invertible. Our choice of domains for these functions depend upon including the origin, if possible, and to have the function be increasing on the intervals of its domain.

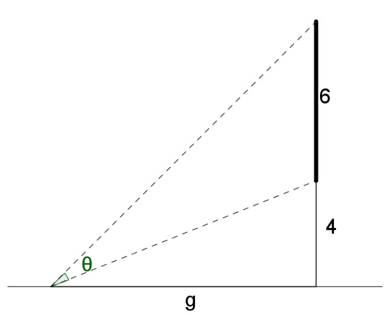
$$\tan : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to (-\infty, \infty) \qquad \tan^{-1} : (-\infty, \infty) \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$$
$$\cot : (0, \pi) \to (-\infty, \infty) \qquad \cot^{-1} : (-\infty, \infty) \to (0, \pi)$$
$$\sec : \left[0, \frac{pi}{2}\right) \cup \left(\frac{\pi}{2}, \pi\right] \to (-\infty, 1] \cup [1, \infty) \qquad \sec^{-1} : (-\infty, 1] \cup [1, \infty) \to \left[0, \frac{pi}{2}\right) \cup \left(\frac{\pi}{2}, \pi\right]$$
$$\csc : \left[-\frac{\pi}{2}, 0\right) \cup \left(0, \frac{\pi}{2}\right] \to (-\infty, 1] \cup [1, \infty) \qquad \csc^{-1} : (-\infty, 1] \cup [1, \infty) \to \left[-\frac{\pi}{2}, 0\right) \cup \left(0, \frac{\pi}{2}\right]$$

Related Content Standards

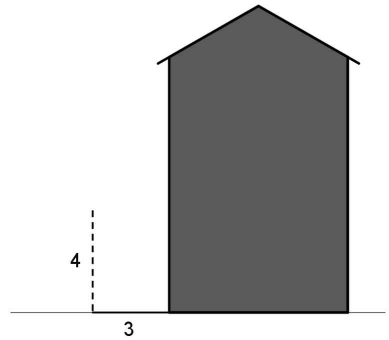
- (HSF.TF.6) Understand that restricting a trigonometric function to a domain on which it is always increasing or always decreasing allows its inverse to be constructed.
- (HSF.TF.7) Use inverse functions to solve trigonometric equations that arise in modeling contexts; evaluate the solutions using technology, and interpret them in terms of the context.

8.8.5 Exercises

1. A billboard with a height of 6 feet is mounted on the side of a building with its bottom edge at a 4 feet above the street. At what distance g along the ground should an observer stand from the building to get the best view?



2. Find the shortest length of a ladder that can be placed against the wall of the house above the 4 foot fence that is placed 3 feet from the house.



8.9 Combinations of Functions

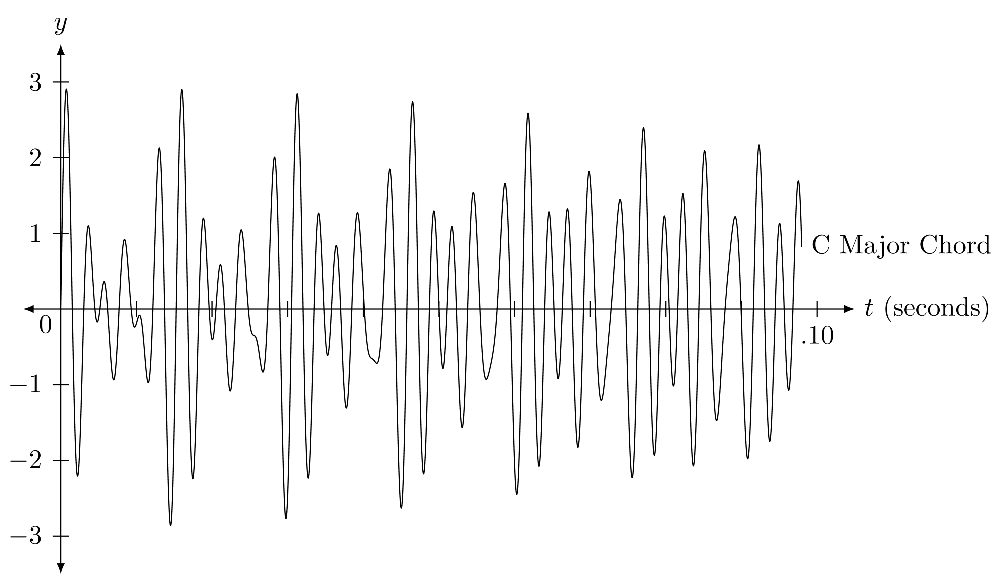
Now that we have an understanding of the main functions in the K-12 curriculum we will look at some of the ways that these functions are combined to create new functions or to model various phenomena.

8.9.1 Addition, Subtraction, Multiplication, and Division of Functions

We will explore how functions can be combined through addition and subtraction by looking at an example from music. The note of middle C has a frequency of 261.63 Hz (cycles per second) and can be modeled by the function $C(t) = \sin(261.63 \cdot (2\pi t))$. If one doubles the number of cycles per second we obtain the next octave higher of the note of C, $\tilde{C}(t) = \sin(523.25 \cdot (2\pi t))$. If we play these notes together, we have

$$C(t) + \tilde{C}(t) = \sin(261.63 \cdot (2\pi t)) + \sin(523.25 \cdot (2\pi t))$$

When the octave is divided into 8 parts, one finds that playing the 1st, 3rd, and 5th notes of the scale creates a very pleasing sound. We can see from the graph of the sum of the three sine functions corresponding the these three notes that the chord has a very complex structure.



The use of noise canceling headphones is another example of the adding and subtracting of functions. In this application, there is a microphone obtaining the noise that is in the surrounding environment and then subtracting that signal from the sound the person is hearing.

$$\text{Sound produced by the headphones} = \text{Sound of the music or speaking desired} \\ - \text{Sound of surrounding environment}$$

There are many additional applications that involve the addition and subtraction of functions including

$$\text{Net Profit} = \text{Revenue} - \text{Expenses}.$$

Some applications that involve the multiplication and/or division of functions include

$$\begin{aligned} \text{Work(time)} &= \text{Force(time)} \cdot \text{Distance(time)} \\ \text{Pressure} &= \frac{\text{Force}}{\text{Area}} \\ \text{Surface Area of a Cylinder} &= 2\pi(\text{radius}) \cdot (\text{height}) + 2\pi(\text{radius})^2 \end{aligned}$$

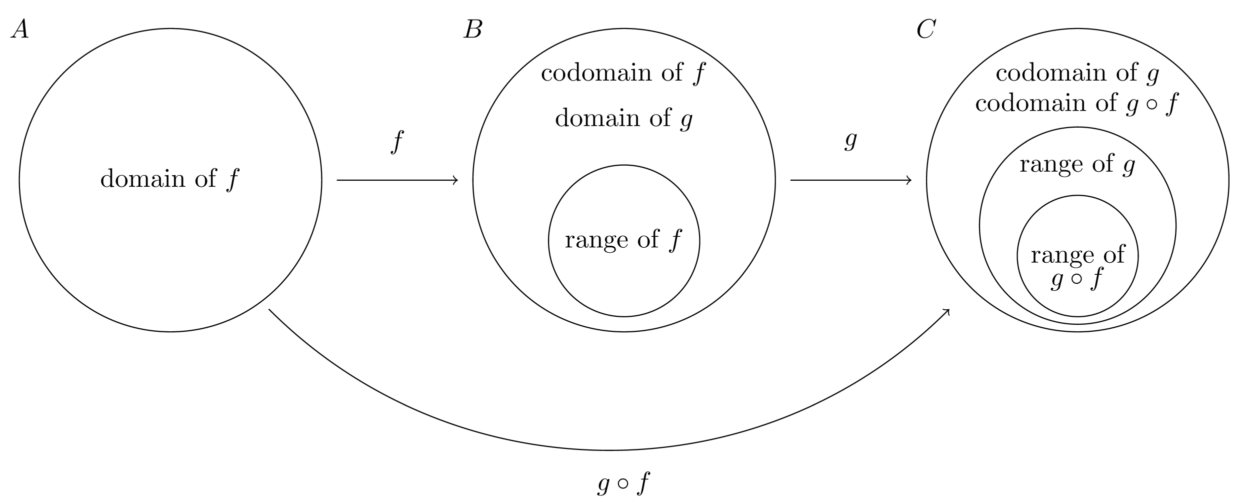
When functions are combined using the basic operations of addition, subtraction, multiplication, and division, the individual functions must have the same type of domain and the domain of the combined function would be the intersection of the domains of the individual functions, excluding any values for which a denominator may be undefined.

8.9.2 Composition of Functions

Another common way to combine functions is through function composition. An example of such a model is finding the area covered by a circular oil slick when the radius of the oil slick is changing with time.

$$\text{Area(time)} = \pi \cdot (\text{radius(time)})^2$$

Recall from Chapter 5 that with the composition of functions, for most cases, the domain of the second function corresponds with the domain of the first function.

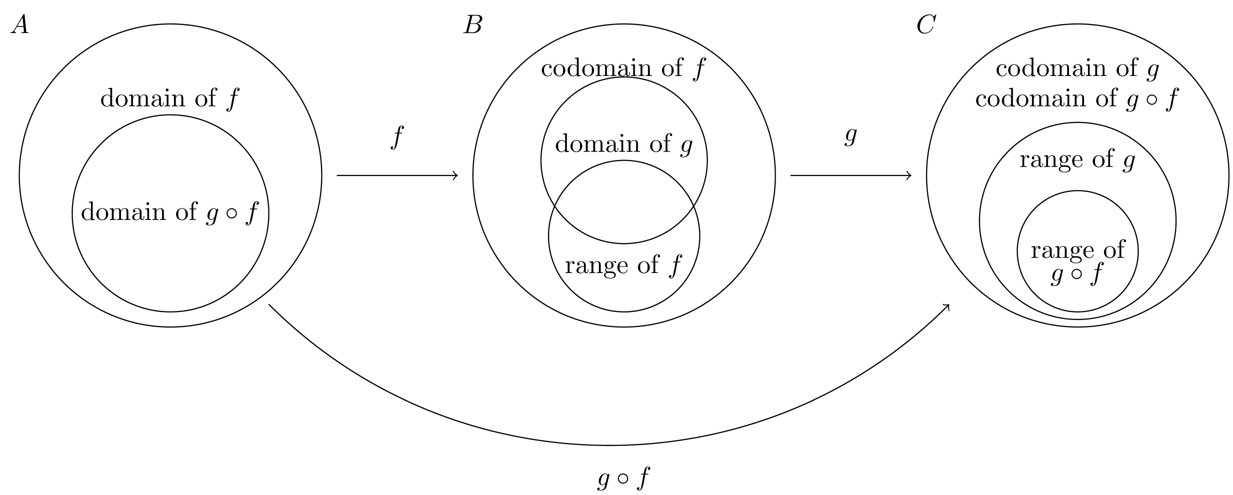


However, we sometimes have compositions of functions for which the domain of the second function is a subset of the co-domain of the first function. For example,

$$g(x) = \sqrt{x}, \quad f(x) = x - 1, \quad \text{and} \quad (g \circ f)(x) = \sqrt{x-1}.$$

In this case, the domain of g is $[0, \infty) \subseteq \mathbb{R}$ and so the domain of the composition of the two functions, $g \circ f$, is $[1, \infty)$.

In order to determine the domain and range of a composition of functions where the domain of the second function does not correspond to the co-domain of the first function, we need to define the pre-image of a function.



If $\phi : U \to V$ and $W \subseteq V$, we define the pre-image of W under the function ϕ as,

$$\phi^{-1}(W) = \{a \in U | \phi(a) \in W\}.$$

Then if

$$f : A \to B \quad \text{and} \quad g : D \subseteq B \to C$$

are functions, the domain of $g \circ f$ is

$$f^{-1}\left((\text{Domain of } g) \cap (\text{Range of } f)\right).$$

Let's look at how this applies to an example of compositions of linear fractional transformations.

Example 8.4. Let $f : \mathbb{R} \setminus \{3\} \to \mathbb{R}$ and $g : \mathbb{R} \setminus \{-1\} \to \mathbb{R}$ be defined by

$$f(x) = \frac{2x - 1}{x - 3} \quad \text{and} \quad g(x) = \frac{5x + 2}{x + 1}.$$

Then the range of f is $\mathbb{R} \setminus \{2\}$ and the domain of g is $\mathbb{R} \setminus \{-1\}$. So the intersection is $\mathbb{R} \setminus \{-1, 2\}$ and

$$f^{-1}\left(\mathbb{R} \setminus \{-1, 2\}\right) = \mathbb{R} \setminus \left\{\frac{4}{3}, 3\right\},$$

since $f\left(\frac{4}{3}\right) = -1$.

So the domain of $g \circ f$ is $\mathbb{R} \setminus \left\{\frac{4}{3}, 3\right\}$ and we can determine the range of $g \circ f$ by removing $g(2)$ from the range of g. So

$$g \circ f : \mathbb{R} \setminus \left\{\frac{4}{3}, 3\right\} \to \mathbb{R} \setminus \{4, 5\}, \quad (g \circ f)(x) = \frac{12x - 11}{3x - 4}$$

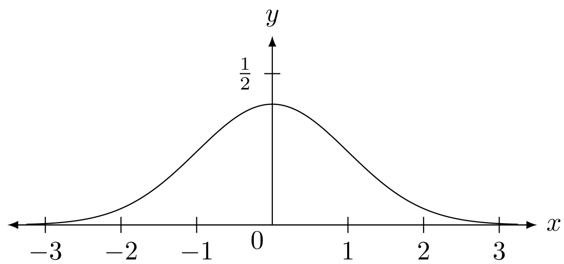
is a surjection.

Once we have determined the domain and range of the composition of functions, we can use properties of each of the individual functions to determine properties of the composition of the functions.

Example 8.5. The function

$$f(x) = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$$

gives the standard normal distribution. If we let $h(x) = -\frac{x^2}{2}$ and $g(x) = \frac{e^x}{\sqrt{2\pi}}$, we see that $f = g \circ h$. Since h is an even function, f must be an even function and so we will have symmetry about the vertical axis. We also know that h is decreasing on $[0\infty)$ and g is always increasing. So $f = g \circ h$ must be decreasing on $[0, \infty)$, and since $\lim_{x\to\infty} h(x) = -\infty$ and $\lim_{x\to-\infty} g(x) = 0$ we know that $\lim_{x\to\infty} f(x) = 0$. Knowing that $f(0) = \frac{1}{\sqrt{2\pi}}$ we have enough knowledge about the function to sketch a graph of the function.



8.9.3 Systems of Equations

The final method of combining multiple functions that we will study are determining the values for which a set of functions have the same input and output. The most common type of problems for which this occurs are systems of linear equations.

Sometimes it is valuable to know the values of x and y for which two equations,

$$\begin{aligned} ax + by &= \alpha \\ cx + dy &= \beta, \end{aligned}$$

are true. In terms of functions, this can be viewed in multiple ways. One option is to think of the two equations as two different linear functions, $f(x) = -\frac{a}{b}x + \frac{\alpha}{b}$ and $g(x) = -\frac{c}{d}x + \frac{\beta}{d}$. We are then determining the values of x for which $f(x) = g(x)$. With this perspective, one can think of this graphically as the points (x, y) for which the two graphs, $y = f(x)$ and $y = g(x)$, intersect. Another method is to view the situation algebraically and determine the values of x for which $-\frac{a}{b}x + \frac{\alpha}{b} = -\frac{c}{d}x + \frac{\beta}{d}$.

Another way to view this situation is through a linear algebra lens where we can write the system of equations in terms of a single function,

$$L\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -\alpha \\ -\beta \end{pmatrix}.$$

We are then trying to find the values for which $L(x, y) = (0, 0)$. These are the values

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

Each of these perspectives has strengths and weaknesses when generalized to different situations. If we generalize to a larger system of equations for which the variables are all linear, $a_1x_1 + a_2x_2 + \cdot a_nx_n = b_n$, then the methods of linear algebra are often most efficient. If we generalize the system of two linear equations to different types of equations with two variables,

$$\begin{matrix} 3x + 2y = 3 \\ x^+y^2 = 4 \end{matrix} \quad \text{or} \quad \begin{matrix} x + y = 8 \\ 3 \cdot 2^x = 0 \end{matrix}$$

then graphical methods are often the more efficient. If we combine these ideas to systems of non-linear equations with more than two variables, numerical analysis techniques are often employed to find numerical approximations to the system of equations.

Related Content Standards

- (8.EE.8) Analyze and solve pairs of simultaneous linear equations.
  a. Understand that solutions to a system of two linear equations in two variables correspond to points of intersection of their graphs, because points of intersection satisfy both equations simultaneously.
  b. Solve systems of two linear equations in two variables algebraically, and estimate solutions by graphing the equations. Solve simple cases by inspection. a.Solve real-world and mathematical problems leading to two linear equations in two variables.
- (HSA.REI.11) Explain why the x-coordinates of the points where the graphs of the equations $y = f(x)$ and $y = g(x)$ intersect are the solutions of the equation $f(x) = g(x)$; find the solutions approximately, e.g., using technology to graph the functions, make tables of values, or find successive approximations. Include cases where $f(x)$ and/or $g(x)$ are linear, polynomial, rational, absolute value, exponential, and logarithmic functions.

8.9.4 Exercises

1. Domain and Range.[2]

   a. Find the domain and range of the following functions.
      - $f(x) = \dfrac{x^2}{(x^2+1)(x+3)}$
      - $f(x) = \sqrt{x^2-16}$
      - $f(x) = \dfrac{1}{\sqrt{2x-5}}$
      - $f(x) = \dfrac{x^2-9}{x\sqrt{x^2-3x+2}}$

   b. Find functions with each the following domains.
      - The set of all real numbers less than or equal to 6
      - The set of all real numbers between -2 and 2 (including both -2 and 2)
      - The set of all real numbers between 1 and 4 (including both 1 and 4).

2. Give an algebraic representation for a function with the indicated domain and range.

   a. domain of $\mathbb{R}$ and range of $\{y \in \mathbb{R} | a \leq y \leq b\}$.
   b. domain of $\{x \in \mathbb{R} | x > 2\}$ and range of $\{y \in \mathbb{R} | y > 1\}$.

3. Find all solutions of the equation
$$x^2 - 3 - \frac{1}{x^2-3} = 0.$$

4. Determine the zeros, vertical asymptotes, and end behavior of the function define by the given algebraic expression.

   a. $f(x) = 2 \cdot 3^{x+2}$
   b. $g(y) = \frac{y^2+1}{y^2-1}$

[2]From *The World of Functions*, part of the Interactive Mathematics Program

5. For each of the following pairs of functions, f and g, give the domain, range, and a simplified formula for $f \circ g$ and for $g \circ f$.

   a. $f(x) = \cos(x)$ and $g(x) = \sin^{-1}(x)$
   b. $f(x) = \frac{1}{x}$ and $g(x) = \sqrt{x^2 - 3x + 2}$
   c. $f(x) = \ln(x-3)$ and $g(x) = x^2 + 4$
   d. $f(x) = \frac{1}{x-2}$ and $g(x) = \frac{1}{x-3}$

6. Find an algebraic representation of a single function, $f(x)$ that has all of the following characteristics.

   - The domain of f is $(-\infty, -1) \cup (-1, 4) \cup (4, \infty)$
   - The range of f is $(-\infty, \infty)$
   - $\lim\limits_{x \to -1^-} f(x) = \infty$
   - $\lim\limits_{x \to -1^+} f(x) = -\infty$
   - $\lim\limits_{x \to 4^-} f(x) = -\infty$
   - $\lim\limits_{x \to 4^+} f(x) = \infty$
   - $f(1) = 0$ and near $x = 1$, $f(x) \approx \dfrac{3}{10}(x-1)$
   - $f(-2) = 0$ and near $x = -2$, $f(x) \approx \dfrac{1}{4}(x+2)^2$
   - $f(3) = 0$ and near $x = 3$, $f(x) \approx \dfrac{-5}{4}(x-3)$
   - As $x \to \pm\infty$, $f(x) \approx \dfrac{x^2 + 3x + 4}{10}$

7. Consider the functions

$$f(x) = 1 + \sqrt{7 - 2x} \quad \text{and} \quad g(x) = \frac{1}{3 - \sqrt{x+5}}.$$

   Let $h(x) = f(x) \cdot g(x)$. What is the domain and range of h? (Round any answers to the nearest tenth.)

8. If $f(x) = |x - 1| - 2$, what is the vertex of $y = f(x+2) - 1$?

9. Let $f(x) = 3\sqrt[4]{x} - \sqrt{x} - 2$. Describe all of the features of the graph of $y = f(x)$ without using technology and then verify with graphing technology.

Part III

GEOMETRY

Chapter 9

Axiomatic Geometry

While the word geometry is derived from ancient Greek words meaning to measure the Earth, the mathematical subject of geometry involves a mixture of theoretical structures and applicable ideas. Egyptian and Babylonian geometry primarily derived from observation, experimentation, and some deductive reasoning to discover relationships between the diameter and circumference of a circle, the Pythagorean theorem, and basic trigonometric relationships. However, the ancient Greeks blended this geometry of measurement centered on applications with their love for logic and reasoning to create the axiomatic system of geometry found in Euclid's *Elements*. The foundational idea of the *Elements* involves the development of a logical system built upon basic accepted elements of definitions, common notions, and postulates.

9.1 Definitions

Cultures define words through the agreement about meanings of the words. In most languages words correspond to multiple definitions or ideas that often depend upon the context and situation to derive the meaning implied. The creation of new definitions and new words often arises from the usage of the words in new ways. For example, the definition of the word 'Google' includes the search for something on the internet, perhaps without even using the search engine created by the company with that name [Edwards and Ward, 2008].

In the language of mathematics, such language is too vague for logical reasoning and determining the truth value of a statement. Mathematical definitions are much more precise and have only one meaning that is usually stipulated at the beginning of the mathematical argument in order to make sure that the one communicating and the one receiving the communication have a well-agreed upon definition. This precision of language is one of the core components of the sixth Standard for Mathematical Practice.

Related Content Standards

- (SMP.6) Attend to precision.

Mathematically proficient students try to communicate precisely to others. They try to use clear definitions in discussion with others and in their own reasoning. They state the meaning of the symbols they choose, including using the equal sign consistently and appropriately. They are careful about specifying units of measure, and labeling axes to clarify the correspondence with quantities in a problem. They calculate accurately and efficiently, express numerical answers with a degree of precision appropriate for the problem context. In the elementary grades, students give carefully formulated explanations to each other. By the time they reach high school they have learned to examine claims and make explicit use of definitions.

Though mathematical definitions are precise, they are not always agreed upon among the general mathematical community. Consider the following two definitions of a circle.

1. A **circle** is a plane figure contained by one [*curved*] line such that all the straight lines falling upon it from one point among those lying within the figure are equal to one another.
2. A **circle** is the set of points in the plane that are equidistant from a single fixed point.

Both definitions of a circle are equally valid, but they refer to two very different objects. The second definition defines the circle as the curve, in which case the circle has no area. The first definition defines the circle as the part of the plane contained within the curve. This illustrates the importance of defining the terms well at the beginning of the argument to make sure that all parties agree upon the definition referenced in the argument.

Related Content Standards

- (HSG.CO.1) Know precise definitions of angle, circle, perpendicular line, parallel line, and line segment, based on the undefined notions of point, line, distance along a line, and distance around a circular arc.

9.1.1 Criteria for Definitions

In the creation of mathematical definitions, Van Dormolen and Zaslavsky [2003] list four criteria as essential and three criteria as preferential.

- Hierarchy - When defining new mathematical objects one should take care to not use mathematical objects that have not been previously defined. For example, one cannot define a right angle without first defining an angle.

- Existence - While one can define a mathematical object using previously defined objects, unless we can also show the existence of such an object, the mathematical logic and proofs regarding that object are irrelevant. For example, we proved the existence of our various number systems by construction in Chapter 4 and we will begin our study of geometry by proving that many geometric objects exist by constructing such objects using a straightedge and compass.

- Equivalence - Some mathematical objects can have multiple definitions equivalent to one another. For instance, one could define an equilateral triangle as a triangle with three equal sides or as a triangle with three equal angles. However, in order to build a consistent system, one of these must be chosen as the definition and the other as a theorem. Then the theorem must be proven to be equivalent to the definition. Then either could be used to prove further properties of equilateral triangles.

- Axiomatization - The final necessary property of a mathematical definition is that it must be part of a larger axiomatic system. Definitions involving number systems are usually a part of an axiomatic system built upon set theory. Definitions in geometry are usually within an axiomatic geometric system similar to that of Euclid. In order to have a hierarchy of definitions, there must be a set of foundational definitions. Such foundational definitions are the axioms of the system.

While lacking any of the above criteria could result in logical contradictions or fallacies, the next three criteria are a part of the beauty and elegance of mathematics.

- Minimality - One could define a square as being a quadrilateral with four equal sides, four right angles, two pairs of parallel sides, and whose diagonals bisect one another at right angles. However, many components of the definition are a direct result of other components. Thus a desirable property in mathematics is to reduce the definition to the minimal amount of requirements in order to better understand the mathematical structures related to the mathematical objects.

- Elegance - When two equivalent definitions exist for a mathematical object, one must make the decision as to which to make the definition and which to make a theorem. Generally, the one chosen as the definition for the object is the one that looks nicer, has fewer words or symbols, and uses more basic mathematical concepts. Such a decision is often subjective as elegance is often in the eye of the beholder.

- Degenerations - Van Dormolen and Zaslavsky [2003] describe this very well with the following example.

  Definition 9.1. A **quadrilateral** is a set of four points A, B, C, D of which no three are collinear and four segments AB, BC, CD, DA.

  Each of the examples in Figure 9.1 satisfies the condition of the definition.

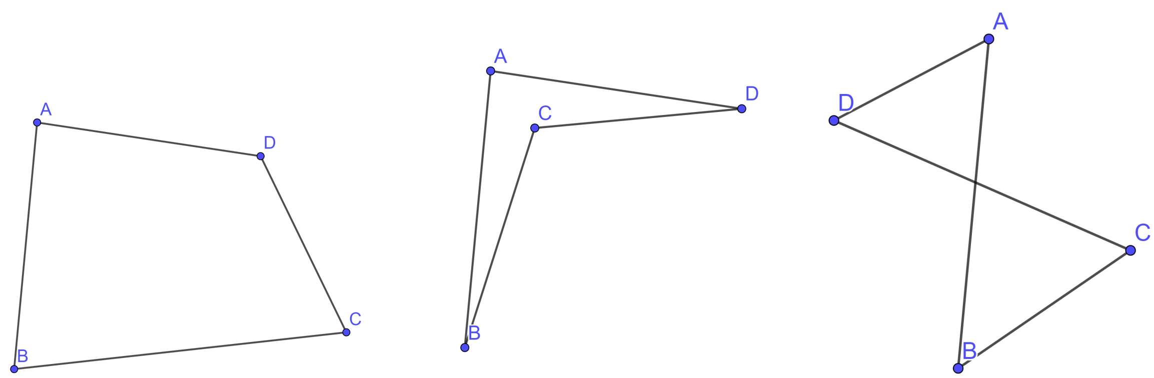


Figure 9.1: Examples of Quadrilaterals

  However, one has the intuition that the third case should not be included and is a degeneration of what is intuitively considered a quadrilateral. As such, such a definition of quadrilaterals would be rejected and replaced with a definition that does not include such degenerations. The most common method of removing these types of quadrilaterals is to require the sides of the quadrilateral to not be self-intersecting.

9.1.2 Concept Image

While the mathematical definition of an object or concept is essential to precise communication and the axiomatization of mathematical structures and systems, less formal understandings of the mathematical concept are essential. Visual representations, mental pictures, and experiences associated with a mathematical concept combine to create a **concept image**. When one thinks about a square, one often does not first think about the precise mathematical definition of a square, but instead has many examples of squares (usually visual) in their mind, along with various properties of squares learned through prior experiences [Vinner, 1991]. Such concept images "may be incomplete or mathematically incorrect, and can include na:{i}ve, non-mathematical associations with the concept name" [Edwards and Ward, 2008, p. 224].

When presenting mathematical tasks to students our ideal is that the students draw upon a mixture of concept definitions and concept images in order to interact with the task as seen in Figure 9.2. However, most often students default is to only address the task through their concept images.

9.1.3 Classification of Two-Dimensional Figures

A good example in the K-12 curriculum where the precision of definitions arise is in the classification of two-dimensional planar figures.

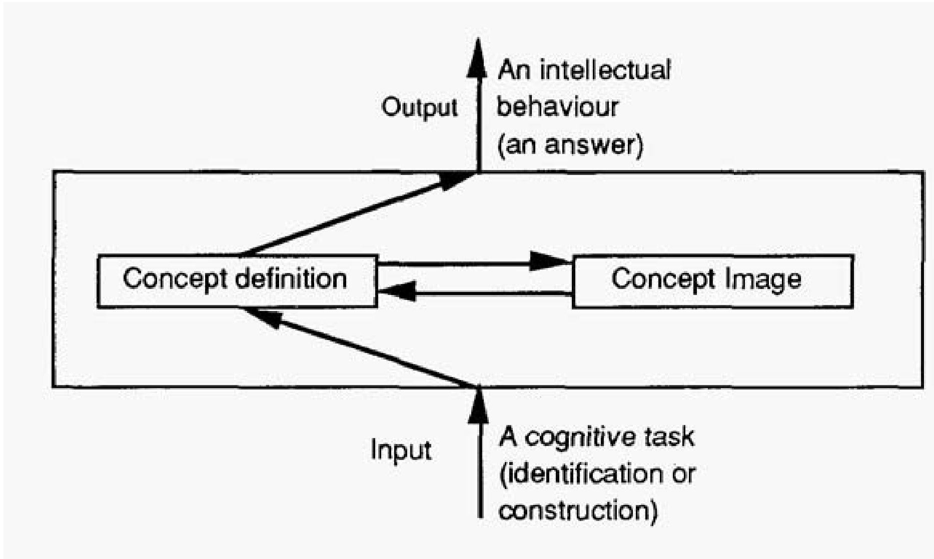


Figure 9.2: Interplay between definition and image [@Vinner1991,p. 71]

Related Content Standards

- (3.G.1) Understand that shapes in different categories (e.g., rhombuses, rectangles, and others) may share attributes (e.g., having four sides), and that the shared attributes can define a larger category (e.g., quadrilaterals). Recognize rhombuses, rectangles, and squares as examples of quadrilaterals, and draw examples of quadrilaterals that do not belong to any of these subcategories.
- (4.G.2) Classify two-dimensional figures based on the presence or absence of parallel or perpendicular lines, or the presence or absence of angles of a specified size. Recognize right triangles as a category, and identify right triangles.
- (5.G.3) Understand that attributes belonging to a category of two-dimensional figures also belong to all subcategories of that category. For example, all rectangles have four right angles and squares are rectangles, so all squares have four right angles.
- (5.G.4) Classify two-dimensional figures in a hierarchy based on properties.

As we see below, the definitions of the two-dimensional figures referenced in these standards are not standardized.

9.1.3.1 Rhombus Definitions

- **Euclid:** Of quadrilateral figures, . . . a rhombus that which is equilateral but not right-angled.
- **Illustrative Mathematics:** Rhombus: A parallelogram with 4 sides with equal length. (A parallelogram is defined to be a quadrilateral with 2 pairs of parallel sides.)
- **Eureka Math/EngageNY:** A rhombus is a quadrilateral with all sides of equal length.

We can see from these definitions that the Euclidean definition excludes squares from being rhombi, while the other two definitions include squares. We can also see that the Illustrative Mathematics definition is not as minimal as the Eureka Math/EngageNY definition. There may be strong pedagogical reasons for avoiding minimality, but it is important for the teacher to know which properties can be derived from other properties.

For further study in this area, Usiskin and Griffin [2008] give a detailed description of the role of definition in the classification of quadrilaterals.

9.1.4 Definitions from Euclid's Elements

Consider the following definitions from Euclid's Elements [Heath, 1908a, p. 153-154], paying particular attention to the various criteria of definitions listed in this section. As new concepts are defined, compare those definitions to your current concept images.

1. A **point** is that which has no part.
2. A **line** is a breadthless length.
3. The extremities of a line are points.
4. A **straight line** is a line which lies evenly with the points on itself.
5. A **surface** is that which has length and breadth only.
6. The extremities of a surface are lines.
7. A **plane surface** is a surface which lies evenly with the straight lines on itself.
8. A **plane angle** is the inclination to one another of two lines in a plane which meet one another and do not lie in a straight line.
9. And when the lines containing the angle are straight, the angle is called **rectilineal**.
10. When a straight line set up on a straight line makes the adjacent angles equal to one another, each of the equal angles is **right**, and the straight line standing on the other is called a **perpendicular** to that on which it stands.
11. An **obtuse angle** is an angle greater than a right angle.
12. An **acute angle** is an angle less than a right angle.
13. A **boundary** is that which is an extremity of anything.
14. A **figure** is that which is contained by any boundary or boundaries.
15. A **circle** is a plane figure contained by one line such that all the straight lines falling upon it from one point among those lying within the figure are equal to one another;
16. And the point is called the **centre** of the circle.
17. A **diameter** of the circle is any straight line drawn through the centre and terminated in both directions by the circumference of the circle, and such a straight line also bisects the circle.
18. A **semicircle** is the figure contained by the diameter and the circumference cut off by it. And the centre of the semicircle is the same as that of the circle.
19. **Rectilineal figures** are those which are contained by straight lines, **trilateral figures** being those contained by three, **quadrilateral** those contained by four, and **multilateral** those contained by more than four straight lines.
20. Of trilateral figures, an **equilateral triangle** is that which has its three sides equal, an **isosceles triangle** that which has two of it sides alone equal, and a **scalene triangle** that which has its three sides unequal.
21. Further, of trilateral figures, a **right-angled triangle** is that which has a right angle, and **obtuse-angled triangle** that which has an obtuse angle, and an **acute-angled triangle** that which has its three angles acute.
22. Of quadrilateral figures, a **square** is that which is both equilateral and right-angled; an **oblong** that which is right-angled but not equilateral; a **rhombus** that which is equilateral but not right-angled; and a **rhomboid** that which has its opposite sides and angles equal to one another but is neither equilateral nor right-angled. And let quadrilaterals other than these be called **trapezia**.
23. **Parallel** straight lines are straight lines which, being in the same plane and being produced indefinitely in both directions, do not meet one another in either direction.

9.1.5 Exercises

1. Classify the quadrilateral figures defined in the definitions from Euclid's Elements in a hierarchy and draw several examples of each type of quadrilateral.

2. Consider the following definitions.

- Given two fixed points F_1, F_2 called the foci and a distance $2a$ which is greater than the distance between the foci, the **ellipse** is the set of points P such that the sum of the distances $|PF_1|, |PF_2|$ is equal to $2a$.
- Given a fixed point F_3 called the center and a distance $r > 0$, the **circle** is the set of points P such that the distance $|PF_3|$ is equal to r.

  According to these definitions, is a circle an ellipse? (Explain your reasoning)

9.2 Axiomatic Systems

One of the primary goals of secondary mathematics education is to develop ways of thinking that extend beyond mathematics. One of these ways of thinking is the ability to use logical arguments to prove mathematical statements using only the most basic of assumptions and how the changing of assumptions can drastically change later outcomes.

An **axiomatic system** consists of certain undefined terms and a list of axioms or postulates concerning these undefined terms. One can then build a mathematical theory by proving propositions, lemmas, theorems, and corollaries using only the axioms, postulates, or previously proven statements. In the process of building the theory, additional definitions are often developed to aid in the precision of language.

In chapters 2, 3, and 4, we worked through an axiomatic system based on the ZFC axioms of set theory to construct the various number systems used in K-12 mathematics. We now turn our attention to axiomatic systems used to study Geometry.

9.2.1 Euclid's Common Notions and Postulates

One of the earliest axiomatic systems was developed in Ancient Greece by Euclid, around 300 B.C., based upon the work of many previous philosophers and mathematicians [Heath, 1908a]. Euclid's system begins with five common notions that are independent from geometry, followed by five postulates of geometry [Heath, 1908a, p. 154-155]. Both the common notions and postulates would be considered axioms in today's language.

Common Notions

The Common Notions are ideas that Euclid determines are generally well accepted.

1. Things which are equal to the same thing are also equal to one another.
2. If equals be added to equals, the wholes are equal.
3. If equals be subtracted from equals, the remainders are equal.
4. Things which coincide with one another are equal to one another.
5. The whole is greater than the part.

We can note that these common notions have a strong connection to the concept of equivalence that we studied in Chapter 3 and the basics of set theory in Chapter 2. In this regard, we can consider these common notions to be the precursor to our modern ZFC axioms, though not as rigorous.

Postulates

1. To draw a straight line from any point to any point.
2. To produce a finite straight line continuously in a straight line.
3. To describe a circle with any centre and distance.

4. That all right angles are equal to one another.
5. That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than two right angles.

Many mathematicians worked to prove that the fifth postulate was a consequence of the prior four postulates. However, we now know that this postulate is independent of the others and with replacing this postulate with a variation one can derive spherical geometry or hyperbolic geometry.

9.2.2 Hilbert's Axioms

David Hilbert was one of the preeminent mathematicians of the late 19th and early 20th centuries and was a leader in the movement to formalize mathematics as a response to the discovery of inconsistencies and paradoxes in the logical system of mathematics being built upon the set theory of George Boole. One of his first endeavors in this direction was to create a set of axioms that would not contain the problems and inconsistencies of Euclid's Elements [Hilbert, 1910].

Hilbert divided his axioms into five categories [Hilbert, 1910, p. 3-26].

I. Axioms of Connection

1. Two distinct points A and B always completely determine a straight line a. We write $AB = a$ or $BA = a$.

2. Any two distict points of a straight line completely determine that line; that is, if $AB = a$ and $AC = a$, where $B \neq C$, then is also $BC = a$.

3. Three points, A, B, C not situated in the same straight line always completely determine a plane α. We write $ABC = \alpha$. We employ also the expressions: A, B, C, "lie in" a; A, B, C "are points of" α, etc.

4. Any three points A, B, C of a plane α, which do not lie in the same straight line, completely determine that plane.

5. If two points A, B of a straight line a lie in a plane α then every point of (the line) a lies in (the plane) α. In this case we say: "The straight line a lies in the plane α," etc.

6. If two planes α, β have a point A in common, then they have at least a second point B in common.

7. Upon every straight line there exist at least two points, in every plane at least three points not lying in the same straight line, and in space there exist at least four points not lying in a plane.

II. Axioms of Order

1. If A, B, C are points of a straight line and B lies between A and C, then B lies also between C and A.

2. If A and C are two points on a straight line, then there exists at least one point B lying between A and C and at least one point D so situated that C lies between A and D.

3. Of any three points situated on a straight line, there is always one and only one which lies between the other two.

4. Any four points A, B, C, D of a straight line can always be so arranged that B shall lie between A and C and also between A and D, and,furthermore, that C shall lie between A and D and also between B and D.

5. Let A, B, C be three points not lying in the same straight line and let a be a straight line lying in the plane ABC and not passing through any of the points A, B, C. Then if the straight line a passes through a point of the segment AB, it will also pass through either a point of the segment BC or a point of the segment AC.

III. Axiom of Parallels

1. In a plane α there can be drawn through any point A, lying outside of a straight line a, one and only one straight line which does not intersect the line a. This straight line is called the parallel to a through the given point A.

IV. Axioms of Congruence

1. If A, B are two points on a straight line a, and if A' is a point upon the same or another straight line a', then, upon a given side of A' on the straight line a', we can always find one and only one point B' so that the segment AB (or BA) is congruent to the segment $A'B'$. We indicate this relation by writing
$$AB \equiv A'B'$$
. Every segment is congruent to itself; that is, we always have
$$AB \equiv AB.$$
2. If a segment AB is congruent to the segment $A'B'$ and also to the segment $A''B''$, then the segment $A'B'$ is congruent to the segment $A''B''$; that is, if $AB \equiv A'B'$ and $AB \equiv A''B''$, then $A'B' \equiv A''B''$.
3. Let AB and BC be two segments of a straight line a which have no points in common aside from the point B, and, furthermore, let $A'B'$ and $B'C'$ be two segments of the same or of another straight line a' having, likewise, no point other than B' in common. Then, if $AB \equiv A'B'$ and $BC \equiv B'C'$, we have $AC \equiv A'C'$.
4. Let an angle (h, k) be given in the plane α and let a straight line a' be given in a plane α'. Suppose also that, in the plane α', a definite side of the staight line a' be assigned. Denote by h' a half-ray of the straight line a' emanating from a point O' of this line. Then in the plane α' there is one and only one half-ray k' such that the angle (h, k), or (k, h), is congruent to the angle (h', k') and at the same time all interior points of the angle (h', k') lie upon the given side of α'. We express this relation by means of the notation
$$\angle(h, k) \equiv \angle(h', k').$$
Every angle is congruent to itself; that is,
$$\angle(h, k) \equiv \angle(h, k) \quad \text{or} \quad \angle(h, k) \equiv \angle(k, h).$$
5. If the angle (h, k) is congruent to the angle (h', k') and to the angle (h'', k''), then the angle (h', k') is congruent to the angle (h'', k''); that is to say, if $\angle(h, k) \equiv \angle(h', k')$ and $\angle(h, k) \equiv \angle(h'', k'')$, then $\angle(h', k') \equiv \angle(h'', k'')$.
6. If, in the two triangles ABC and $A'B'C'$, the congruences
$$AB \equiv A'B', \quad AC \equiv A'C', \quad \angle BAC \equiv \angle B'A'C'$$
hold, then the congruences
$$\angle ABC \equiv \angle A'B'C' \quad \text{and} \quad \angle ACB \equiv A'C'B'$$
also hold.

V. Axiom of Continuity

1. Let A_1 be any point upon a straight line between the arbitrarily chosen points A and B. Take the points $A_2, A_3, A_4, . . . $ so that A_1 lies between A and A_2, A_2 between A_1 and A_3, A_3 between A_2 and A_4, etc. Moreover, let the segments

$$AA_1, A_1A_2, A_2A_3, A_3A_4, \ldots$$

be equal to one another. Then, among this series of points, there always exists a certain point A_n such that B lies between A and A_n.

9.2.3 School Mathematics Study Group Axioms

Following World War II, a group of mathematicians and mathematics educators called the School Mathematics Study Group (SMSG) developed a curriculum for K-12 education with funding from the National Science Foundation. In order to simplify the axiomatic systems of Euclid and Hilbert they created a system of 22 postulates that is equivalent to those of Euclid and Hilbert. Their primary goal was to not focus on minimality, but on accessibility, by allowing some of the postulates to be consequences of prior postulates. This allowed the curriculum to instead focus on certain steps in the process of building a consistent geometric system.

The School Mathematics Study Group [1960] axiomatic system has three primary undefined terms of point, line, and plane. From these terms we have the following postulates.

1. (Line Uniqueness) Given any two distinct points there is exactly one line that contains them.
2. (Distance Postulate) To every pair of distinct points there corresponds a unique positive number. This number is called the distance between the two points.
3. (Ruler Postulate) The points of a line can be placed in a correspondence with the real numbers such that:
   a. To every point of the line there corresponds exactly one real number.
   b. To every real number there corresponds exactly one point of the line.
   c. The distance between two distinct points is the absolute value of the difference of the corresponding real numbers.
4. (Ruler Placement Postulate) Given two points P and Q of a line, the coordinate system can be chosen in such a way that the coordinate of P is zero and the coordinate of Q is positive.
5. (Existence of Points) Every plane contains at least three non-collinear points. Space contains at least four non-coplanar points.
6. (Points on a Line Lie in a Plane) If two points lie in a plane, then the line containing these points lies in the same plane.
7. (Plane Uniqueness) Any three points lie in at least one plane, and any three non-collinear points lie in exactly one plane.
8. (Plane Intersection) If two planes intersect, then that intersection is a line.
9. (Plane Separation Postulate) Given a line and a plane containing it, the points of the plane that do not lie on the line form two sets such that:
   a. each of the sets is convex;
   b. if P is in one set and Q is in the other, then segment $\overline{PQ}$ intersects the line.
10. (Space Separation Postulate) The points of space that do not lie in a given plane form two sets such that:

a. Each of the sets is convex.
b. If P is in one set and Q is in the other, then segment $\overline{PQ}$ intersects the plane.

11. (Angle Measurement Postulate) To every angle there corresponds a real number between 0° and 180°.

12. (Angle Construction Postulate) Let $\overrightarrow{AB}$ be a ray on the edge of the half-plane H. For every r between 0 and 180, there is exactly one $\overrightarrow{AP}$ with P in H such that $m\angle PAB = r$.

13. (Angle Addition Postulate) If D is a point in the interior of $\angle BAC$, then $m\angle BAC = m\angle BAD + m\angle DAC$.

14. (Supplement Postulate) If two angles form a linear pair, then they are supplementary.

15. (SAS Postulate) Given a one-to-one correspondence between two triangles (or between a triangle and itself). If two sides and the included angle of the first triangle are congruent to the corresponding parts of the second triangle, then the correspondence is a congruence.

16. (Parallel Postulate) Through a given external point there is at most one line parallel to a given line.

17. (Area of Polygonal Region) To every polygonal region there corresponds a unique positive real number called the area.

18. (Area of Congruent Triangles) If two triangles are congruent, then the triangular regions have the same area.

19. (Summation of Areas of Regions) Suppose that the region R is the union of two regions R_1 and R_2. If R_1 and R_2 intersect at most in a finite number of segments and points, then the area of R is the sum of the areas of R_1 and R_2.

20. (Area of a Rectangle) The area of a rectangle is the product of the length of its base and the length of its altitude.

21. (Volume of Rectangular Parallelepiped) The volume of a rectangular parallelepiped is equal to the product of the length of its altitude and the area of its base.

22. (Cavalieri's Principle) Given two solids and a plane. If for every plane that intersects the solids and is parallel to the given plane such that the two intersections determine regions that have the same area, then the two solids have the same volume.

9.2.4 Discussion

While Euclid has significantly fewer axioms, it also has some holes in the logical consequences particularly in the proof of the SAS triangle congruence. It also makes some assumptions regarding properties of the real number system. In particular, it assumes completeness of the real numbers. Hilbert deals with the real number system with the Axiom of Continuity and the SMSG system uses properties of the real numbers built up in parallel with the Distance Postulate and the Ruler Postulate.

While the goal of the formalists to develop a foundation for all of mathematics from a single set of axioms can never be achieved (see the Gödel Incompleteness Theorems), the goal of using axiomatic methods for mathematics education in order to help students to reason with logic and to justify their arguments is attainable.

9.2.5 Exercises

1. Compare the parallel postulates/axioms between the different systems. How are they similar and different?

2. Discuss the similarities of the Common Notions of Euclid with algebra of sets from Chapter 2.

3. The introduction to High School: Geometry in the Common Core state,

   > During high school, students begin to formalize their geometry experiences from elementary and middle school, using more precise definitions and developing careful proofs. Later in college some students develop Euclidean and other geometries carefully from a small set of axioms.

   What is the role of studying geometry from an axiomatic perspective in secondary mathematics?

9.3 Euclid's Basic Constructions

Now that we have established the role of definitions and axioms we will look at how those definitions and axioms are used to prove geometric theorems using Euclid's system. Euclid's geometry is abstract in that one cannot see the objects of a point or a line (as they have no breadth). However, we represent points and lines with symbols that we can draw on a piece of paper. So in the proofs we must remember that the drawings are not the proof and are not the actual points, lines, and circles. They are instead representations of points, lines, and circles and assist in the communication of the mathematical proof.

Figure 9.3: Straightedge and Compass

To assist in the representations we often use a straight edge (usually a ruler ignoring the markings) to represent straight lines and we use a compass (either collapsible or fixed) to describe a circle with a fixed center and radius.

Related Content Standards

- (HSG.CO.12) Make formal geometric constructions with a variety of tools and methods (compass and straightedge, string, reflective devices, paper folding, dynamic geometric software, etc.). *Copying a segment; copying an angle; bisecting a segment; bisecting an angle; constructing perpendicular lines, including the perpendicular bisector of a line segment; and constructing a line parallel to a given line through a point not on the line.*

We can also use dynamic geometry software like Desmos or GeoGebra to create representations of lines and circles.

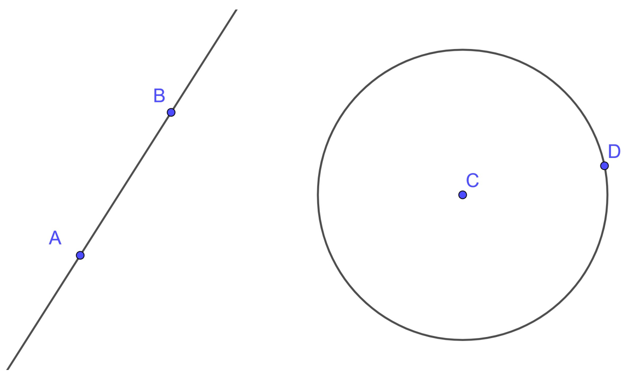


Figure 9.4: Line and Circle Constructed using GeoGebra

As we work through some of the Propositions from Euclid's Elements we will use some of these representations within the proofs in order to communicate the ideas of the proof. For each of the propositions we will give the proposition using the translation of Heath [Heath, 1908a] in order to better understand the historical context. We will then give a more modern version of the proposition.

9.3.1 Initial Propositions

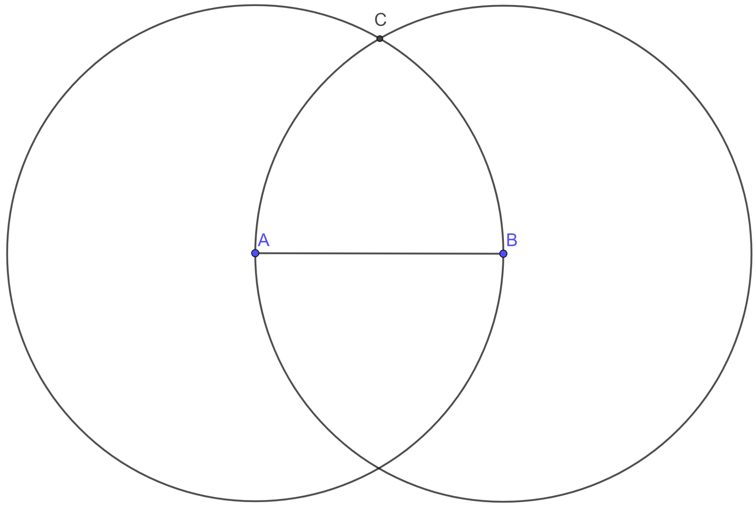
We will include the proofs of some of Euclid's initial propositions below in order to understand the idea of the axiomatic method and to demonstrate some of Euclid's basic constructions. Proofs of the remainder of the propositions can easily be found online.

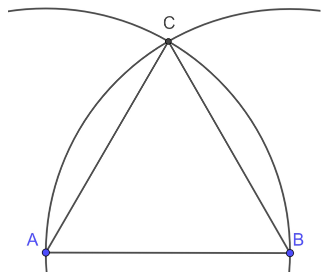
Proposition 9.1 (Euclid's Proposition 1: Construction of Equilateral Triangle)**.** *On a given finite straight line to construct an equilateral triangle.*

This proposition states that if you are given a line segment, you can use a compass and straightedge to construct and equilateral triangle where the line segment is one of the sides.

Proof. We start with a given finite straight line AB, labeled by its endpoints. Such a finite straight line exists by Postulate 1.

Postulate 3 implies that we can describe a circle centered at A through the point B and a second circle centered at B through the point A.





The point C is an intersection of the two circles. The existence of such a point, and that it lies in the same plane as AB are some of the first issues raised with the Euclidean system of axioms that are resolved in the Hilbert and SMSG systems.

Since C lies on the circle centered at A through the point B, AC must be the same length as AB. Since C also lies on the circle centered at B through A, BC must be the same length as BA. Then using Common Notions 1 and 4 we can conclude that the three sides of $\triangle ABC$ have the same length and so form an equilateral triangle. $\square$

Euclid's second and third propositions allow us to construct a segment of the same length as a given segment at another point and along a line through that point.

Proposition 9.2 (Euclid's Proposition 2)**.** *To place at a given point (as an extremity) a straight line equal to a given straight line.*

Proposition 9.3 (Euclid's Proposition 3)**.** *Given two unequal straight lines, to cut off from the greater a straight line equal to the less.*

Euclid's proof of Proposition 4 uses a technique called superposition. In our current language, Euclid considers a translation to maintain distances and angles without proving such. We can note that this proposition is the 15th postulate in the SMSG axioms as it is something that is difficult to prove geometrically without analytic techniques using a coordinate plane.

Proposition 9.4 (Euclid's Proposition 4: SAS Triangle Congruence)**.** *If two triangles have the two sides equal to two sides respectively, and have the angles contained by the equal straight lines equal, they will also have the base equal to the base, the triangle will be equal to the triangle, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend.*

9.3.2 Isosceles Triangles

We can begin using SAS triangle congruence to prove results regarding isosceles triangle. While isosceles triangles are defined by Euclid as triangles with exactly two equal sides, the next two propositions show that these could be defined equivalently by having exactly two equal angles. We first show that two equal sides implies two equal angles.

Proposition 9.5 (Euclid's Proposition 5). *In isosceles triangles the angles at the base are equal to one another, and, if the equal straight lines be produced further, the angles under the base will be equal to one another.*

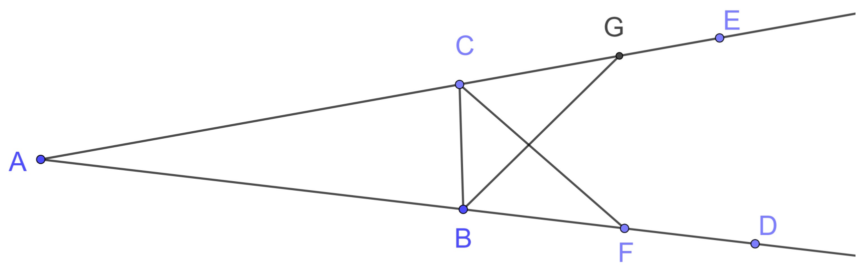
As you read through the proof from Heath's translation of Euclid's Elements [1908a, p. 251], rewrite the proof using modern terminology and verify the truthfulness of each statement based upon the prior Propositions.

Proof. Let $\triangle ABC$ be an isosceles triangle having the side AB equal to the side AC, and let the straight lines BD and CE be produced further in a straight line with AB and AC.

I say that the angle $\angle ABC$ equals the angle $\angle ACB$, and the angle $\angle CBD$ equals the angle $\angle BCE$.

Take an arbitrary point F on BD. Cut off AG from AE the greater equal to AF the less, and join the straight lines FC and GB.

Since AF equals AG, and AB equals AC, therefore the two sides FA and AC equal the two sides GA and AB, respectively, and they contain a common angle, the angle $\angle FAG$.



Therefore the base FC equals the base GB, the triangle $\triangle AFC$ equals the triangle $\triangle AGB$, and the remaining angles equal the remaining angles respectively, namely those opposite the equal sides, that is, the angle $\angle ACF$ equals the angle $\angle ABG$, and the angle $\angle AFC$ equals the angle $\angle AGB$.

Since the whole AF equals the whole AG, and in these AB equals AC, therefore the remainder BF equals the remainder CG.

But FC was also proved equal to GB, therefore the two sides BF and FC equal the two sides CG and GB respectively, and the angle $\angle BFC$ equals the angle $\angle CGB$, while the base BC is common to them. Therefore the triangle $\triangle BFC$ also equals the triangle $\triangle CGB$, and the remaining angles equal the remaining angles respectively, namely those opposite the equal sides. Therefore the angle $\angle FBC$ equals the angle $\angle GCB$, and the angle $\angle BCF$ equals the angle $\angle CBG$.

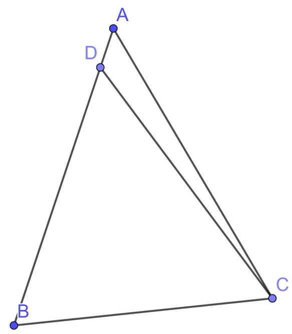
Accordingly, since the whole angle $\angle ABG$ was proved equal to the angle $\angle ACF$, and in these the angle $\angle CBG$ equals the angle $\angle BCF$, the remaining angle $\angle ABC$ equals the remaining angle $\angle ACB$, and they are at the base of the triangle $\triangle ABC$. But the angle $\angle FBC$ was also proved equal to the angle $\angle GCB$, and they are under the base.

Therefore in isosceles triangles the angles at the base equal one another, and, if the equal straight lines are produced further, then the angles under the base equal one another. $\square$

We now need to prove the converse that a triangle with two equal angles also has two equal sides.

Proposition 9.6 (Euclid's Proposition 6). *If in a triangle two angles be equal to one another, the sides which subtend the equal angles will also be equal to one another.*

Proof. Let $\triangle ABC$ be a triangle having the angle $\angle ABC$ equal to the angle $\angle ACB$.



If AB does not equal AC, then one of them is greater. Let AB be greater. Cut off DB from AB the greater equal to AC the less, and join DC.

Since DB equals AC, and BC is common, therefore the two sides DB and BC equal the two sides AC and CB respectively, and the angle $\angle DBC$ equals the angle $\angle ACB$. Therefore the base DC equals the base AB, and the triangle $\triangle DBC$ equals the triangle $\triangle ACB$, the less equals the greater, which is absurd. Therefore AB is not unequal to AC, it therefore equals it.

Therefore if in a triangle two angles equal one another, then the sides opposite the equal angles also equal one another. $\square$

9.3.3 Side-Side-Side Congruence

We will skip the proofs of the next two propositions, as the proofs do not add much to our understanding. However, we will include the statements of the propositions as they will be very useful later.

Proposition 9.7 (Euclid's Proposition 7). *Given two straight lines constructed on a straight line (from its extremities) and meeting in a point, there cannot be constructed on the same straight line (from its extremities), and on the same side of it, two other straight lines meeting in another point and equal to the former two respectively, namely each to that which has the same extremity with it.*

This first Proposition is a little confusing, but it is stating that if one of the sides of a triangle is given, and the lengths of the other two sides is also given, then one can construct two possible triangles with that information. One on each side of the given side. The next proposition uses this to prove that any two triangles that have the same side measures are congruent to each other, since the equal side lengths force equality of the measures of the angles.

Proposition 9.8 (Euclid's Proposition 8: SSS Triangle Congruence). *If two triangles have the two sides equal to two sides respectively, and have also the base equal to the base, they will also have the angles equal which are contained by the equal straight lines.*

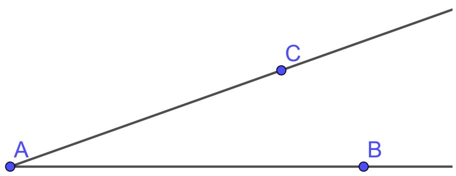
9.3.4 Angles

With these propositions we will look at some of the important constructions regarding angles, points, and sides.

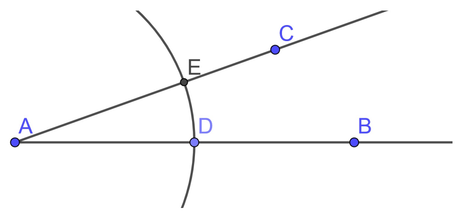
Our first construction bisects angles.

Proposition 9.9 (Euclid's Proposition 9: Construction of Angle Bisector)**.** *To bisect a given rectilinear angle.*

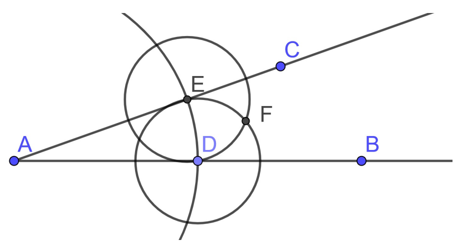
Proof. Let $\angle BAC$ be a given rectilinear angle.



Then we can construct a circle centered at A whose radius is less than either AB or AC. We can then label the point at which the circle intersects AB as D and the point at which the circle intersects AC as E.



We can then construct a circle centered at D through the point E and a circle centered at E through the point D. These two circles intersect at a point that we will label F.



We now focus on the triangles $\triangle ACD$ and $\triangle AEF$.

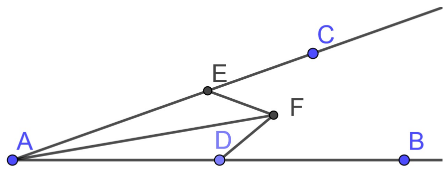
We know that AE and AD are equal because D and E are both on a circle centered at A. Since F is an intersection of the circles centered at D and E, EF is equal to DF. Since AF is equal to AF we have the sides of the two triangles all being equal. By Proposition 8, we have that $\triangle ACD$ and $\triangle AEF$ are congruent and so $\angle EAF$ equals $\angle DAF$. So AF bisects the angle $\angle BAC$ □

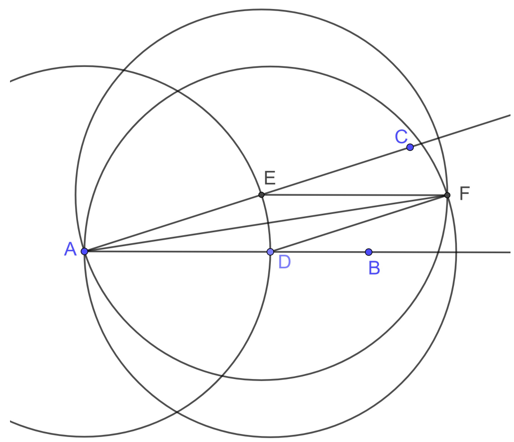
Another option for the construction process is that the circles centered at D and E pass through the point A. The other point of intersection for these circles can be labeled F.

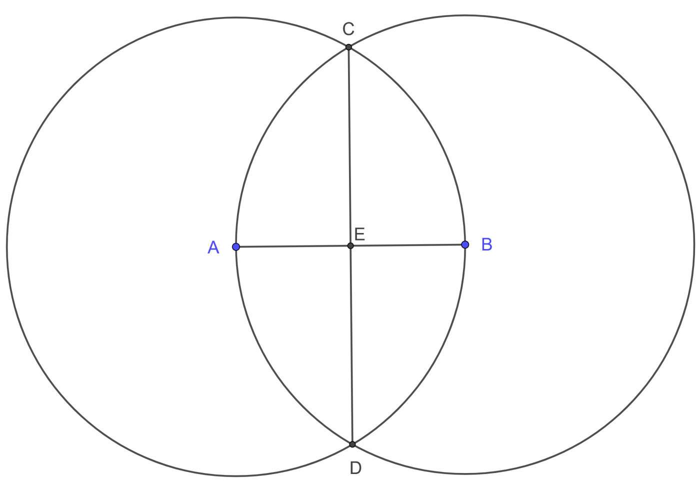
The segment AF bisects the angle $\angle BAC$. We will show later that $AEFD$ forms a rhombus, generating parallel lines among other things.

9.3.5 Perpendicular Lines

The next few propositions involve the construction of perpendicular lines. If we start with a finite straight line AB, we can use the following construction to create the perpendicular bisector. We will leave the details of the proof to the reader to work through.


C
E
F
A
D
B

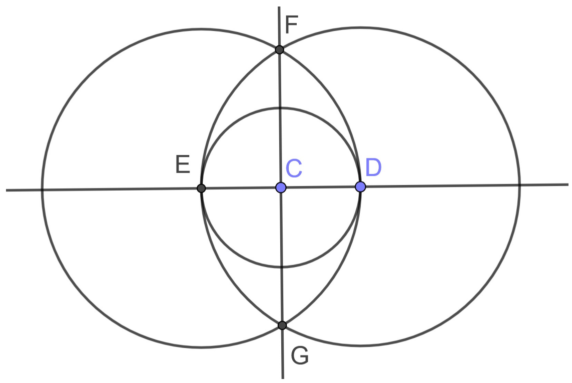

C
E
F
A
D
B


C
A
E
B
D

Proposition 9.10 (Euclid's Proposition 10: Construction of Perpendicular Bisector)**.** *To bisect a given finite straight line.*

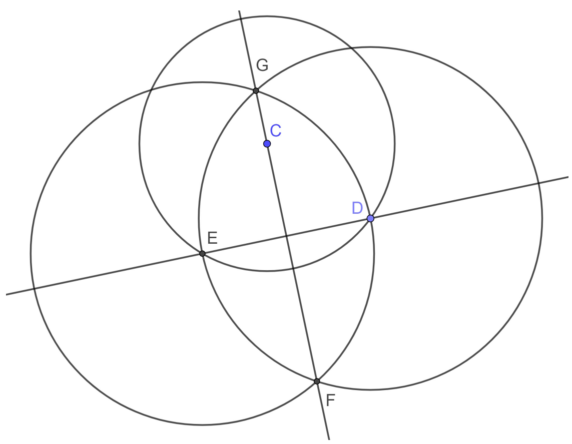
Proposition 9.11 (Euclid's Proposition 11: Construction of a Perpendicular Line)**.** *To draw a straight line at right angles to a given straight line from a given point on it.*

We will not give the details of the proof, but will include the appropriate visual construction that gives the main ideas.



Proposition 9.12 (Euclid's Proposition 12)**.** *To a given infinite straight line, from a given point which is not on it, to draw a perpendicular straight line.*

The proof of this proposition uses the information from the following diagram.



9.3.6 Exercises

1. Discuss Euclid's use of the word 'equal' in the context of the discussion in Chapter 3.
2. Write out the details of the proof of Proposition 9.10.
3. Write out the details of the proof of Proposition 9.11.
4. Write out the details of the proof of Proposition 9.12.
5. Complete the following constructions with a physical compass and straightedge, and with a geometry software (using just the straightedge and circles and the more complex construction tools).

    a. Given a segment of length a, construct a segment of length $2a$.

b. Given a segment of length a and a segment of length b, construct a segment of length $a + b$
c. Construct a triangle with side lengths a, b, and c (with these given).
d. Construct a square with side length a.
e. Construct a rhombus that is not a square with side length a.

9.4 Angles, Parallel Lines, and Parallelograms

In this section we will review some of the properties of angles and parallel lines. Many of these properties are dependent upon congruent triangles so we will first list out these for use later.

Proposition 9.13 (Euclid's Proposition 4: SAS Triangle Congruence)**.** *If two triangles have the two sides equal to two sides respectively, and have the angles contained by the equal straight lines equal, they will also have the base equal to the base, the triangle will be equal to the triangle, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend.*

Proposition 9.14 (Euclid's Proposition 8: SSS Triangle Congruence)**.** *If two triangles have the two sides equal to two sides respectively, and have also the base equal to the base, they will also have the angles equal which are contained by the equal straight lines.*

Proposition 9.15 (Euclid's Proposition 26: ASA and AAS Triangle Congruence)**.** *If two triangles have two angles equal to two angles respectively, and one side equal to one side, namely, either the side adjoining the equal angles, or that opposite one of the equal angles, then the remaining sides equal the remaining sides and the remaining angle equals the remaining angle.*

9.4.1 Angles and Parallel Lines

Related Content Standards

- (HSG.CO.9) Prove theorems about lines and angles. *Theorems include: vertical angles are congruent; when a transversal crosses parallel lines, alternate interior angles are congruent and corresponding angles are congruent; points on a perpendicular bisector

In the proofs of each of the following propositions we will only use propositions with a lower number, combined with Euclid's Common Notions and Postulates.

Proposition 9.16 (Euclid's Proposition 13)**.** *If a straight line set up on a straight line make angles, it will make either two right angles or angles equal to two right angles.*

Proof. Let any straight line AB standing on the straight line CD make the angles $\angle CBA$ and $\angle ABD$.

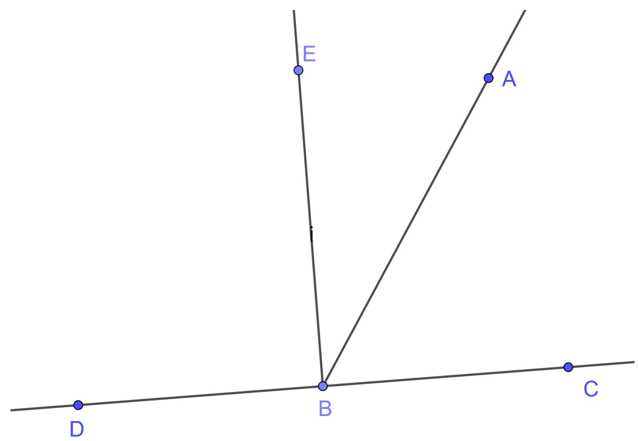
If the angle $\angle CBA$ equals the angle $\angle ABD$, then they are two right angles (Definition 10). If not, draw BE from the point B at right angles to CD. Therefore the angles $\angle CBE$ and $\angle EBD$ are two right angles.

Since the angle $\angle CBE$ equals the sum of the two angles $\angle CBA$ and $\angle ABE$, add the angle $\angle EBD$ to each, therefore the sum of the angles $\angle CBE$ and $\angle EBD$ equals the sum of the three angles $\angle CBA$, $\angle ABE$, and $\angle EBD$.

Using the common notions we can see that the sum of the angles $\angle DBA$ and $\angle ABC$ equals two right angles.

Therefore if a straight line stands on a straight line, then it makes either two right angles or angles whose sum equals two right angles. $\square$

Using this property along with the Common Notions one can also prove its converse.

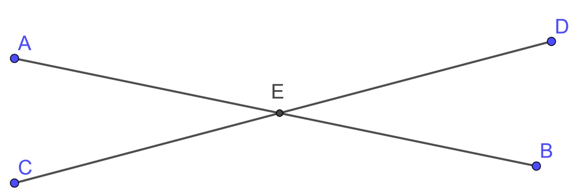


Proposition 9.17 (Euclid's Proposition 14)**.** *If with any straight line, and at a point on it, two straight lines not lying on the same side make the adjacent angles equal to two right angles, two straight lines will be in a straight line with one another.*

We are now able to prove that vertical angles are equal to one another.

Proposition 9.18 (Euclid's Proposition 15)**.** *If two straight lines cut one another, they make the vertical angles equal to one another.*

Proof. Let the straight lines AB and CD cut one another at the point E.



Since the straight line AE stands on the straight line CD making the angles $\angle CEA$ and $\angle AED$, therefore the sum of the angles $\angle CEA$ and $\angle AED$ equals two right angles.

Again, since the straight line DE stands on the straight line AB making the angles $\angle AED$ and $\angle DEB$, therefore the sum of the angles $\angle AED$ and $\angle DEB$ equals two right angles.

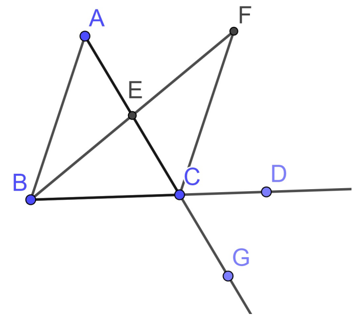
But the sum of the angles $\angle CEA$ and $\angle AED$ was also proved equal to two right angles, therefore the sum of the angles $\angle CEA$ and $\angle AED$ equals the sum of the angles $\angle AED$ and $\angle DEB$. Subtract the angle $\angle AED$ from each. Then the remaining angle $\angle CEA$ equals the remaining angle $\angle DEB$.

Similarly it can be proved that the angles $\angle BEC$ and $\angle AED$ are also equal. □

Now that we have proven that vertical angles are congruent, we will move to a study of angles related to parallel lines and transversals. But we first need to prove a property of interior and exterior angles of a triangle. A stronger result that the exterior angle is equal to the sum of the opposite interior angles is dependent upon the parallel postulate and so it is delayed by Euclid until later.

Proposition 9.19 (Euclid's Proposition 16)**.** *In any triangle, if one of the sides is produced, then the exterior angle is greater than either of the interior and opposite angles.*

Proof. Let $\triangle ABC$ be a triangle, and let one side of it BC be extended to a point D.



Bisect AC at E and create a finite straight line BE and extend it to a point F such that EF is equal to BE. Create the finite straight line FC and extend AC to a point G.

Since AE equals EC, and BE equals EF, and the angle $\angle AEB$ equals the angle $\angle FEC$, since they are vertical angles, the triangle $\triangle ABE$ equals the triangle $\triangle CFE$, and the remaining angles equal the remaining angles respectively, namely those opposite the equal sides. Therefore the angle $\angle BAE$ equals the angle $\angle ECF$.

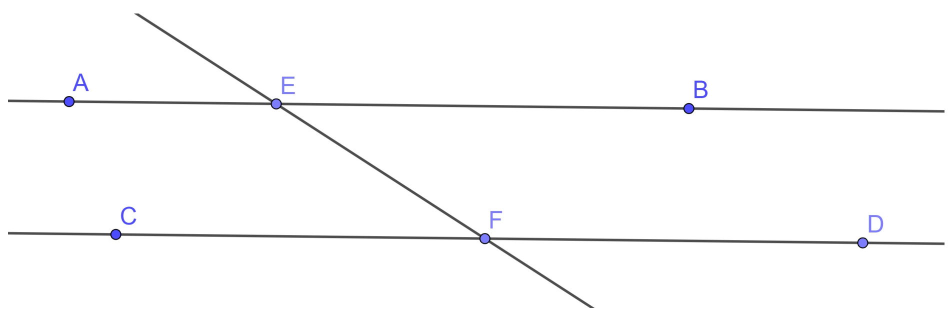
But the angle $\angle ECD$ is greater than the angle $\angle ECF$, therefore the angle $\angle ACD$ is greater than the angle $\angle BAE$.

Similarly, if BC is bisected, then the angle $\angle BCG$, that is, the angle $\angle ACD$, can also be proved to be greater than the angle $\angle ABC$.

Therefore in any triangle, if one of the sides is extended, then the exterior angle is greater than either of the interior and opposite angles. □

Proposition 9.20 (Euclid's Proposition 27)**.** *If a straight line falling on two straight lines makes the alternate angles equal to one another, then the straight lines are parallel to one another.*

Proof. Let the straight line EF falling on the two straight lines AB and CD make the alternate angles $\angle AEF$ and $\angle EFD$ equal to one another.



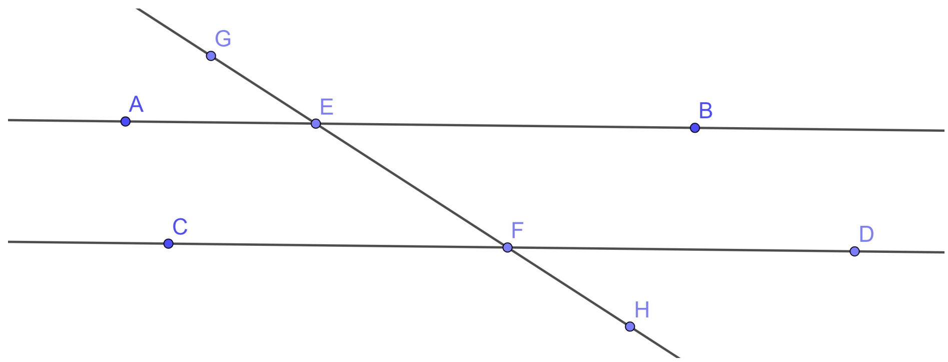
If AB and CD are not parallel, then they meet at a point. We can assume without any loss of generality that they meet at a point in the direction of B and D. We will label the point of intersection as G.

Then, in the triangle $\triangle GEF$, the exterior angle $\angle AEF$ equals the interior and opposite angle $\angle EFG$, which contradicts Proposition 9.19.

Therefore AB is parallel to CD. □

Proposition 9.21 (Euclid's Proposition 28)**.** *If a straight line falling on two straight lines makes the exterior angle equal to the interior and opposite angle on the same side, or the sum of the interior angles on the same side equal to two right angles, then the straight lines are parallel to one another.*

Proof. Let EF be a straight line falling on the two straight lines AB and CD such that the exterior angle $\angle GEA$ is equal to the interior angle $\angle EFC$



If we assume that AB is not parallel to CD, then they intersect at a point J. If J is in the direction of A and C, then we would have the exterior angle $\angle GEA$ being equal to the interior angle $\angle EFC$, contradicting Proposition 9.19. If J is in the direction of B and D, we can use the Common Notions and Proposition 9.16 to prove that angles $\angle GEB$ and $\angle EFD$ are equal to create an exterior angle equal to an interior angle, contradicting Proposition 9.19.

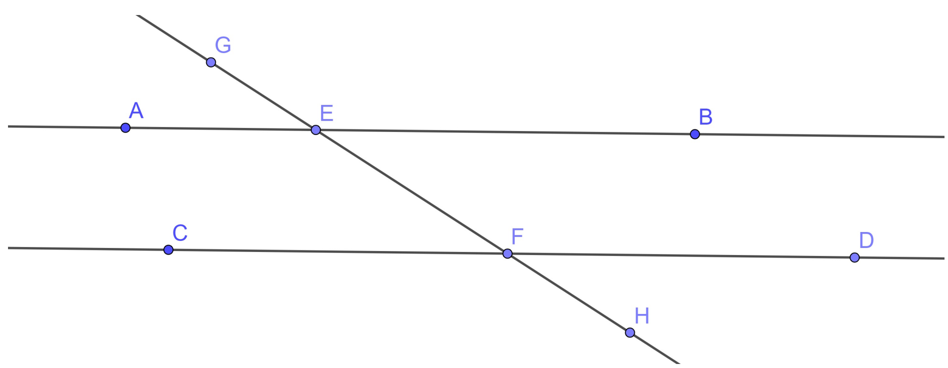
Therefore, AB is parallel to CD.

If instead the sum of the interior angles on the same side are equal to two right angles Proposition 9.16 and the Common Notions will lead us to conclude that the exterior angle is equal to the interior and opposite angle on the same side. This would then imply that AB is parallel to CD. □

Proposition 9.22 is the converse of Propositions 9.20 and 9.21. It is also the first of Euclid's propositions that makes use of the fifth postulate.

Proposition 9.22 (Euclid's Proposition 29)**.** *A straight line falling on parallel straight lines makes the alternate angles equal to one another, the exterior angle equal to the interior and opposite angle, and the sum of the interior angles on the same side equal to two right angles.*

Proof. Let the straight line GH fall on the parallel straight lines AB and CD.



If the angle $\angle AEF$ does not equal the angle $\angle EFD$, then one of them is greater. Let the angle $\angle AEF$ be greater.

Add the angle $\angle BEF$ to each. Therefore the sum of the angles $\angle AEF$ and $\angle BEF$ is greater than the sum of the angles $\angle BEF$ and $\angle EFD$.

But sum of the angles $\angle AEF$ and $\angle BEF$ equals two right angles. Therefore the sum of the angles $\angle BEF$ and $\angle EFD$ is less than two right angles.

Then by Postulate 5, the lines AB and CD are not parallel, contradicting the hypothesis.

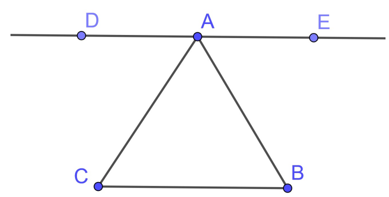
Therefore, the angles $\angle AEF$ and $\angle EFD$ are equal and, by a similar set of arguments, the angles $\angle BEF$ and $\angle EFC$ are equal resulting in the alternate angles equal to one another.

Using Propositions 9.16 and 9.18 we find that the exterior angle equal to the interior and opposite angle, and the sum of the interior angles on the same side equal to two right angles. □

One of the most used results of this is that sum of angles is equal to two right angles.

Theorem 9.1. *The sum of angles of a triangle is equal to two right angles.*

Proof. Let $\triangle ABC$ be given with a line through A parallel to BC.



From Proposition 9.16 and the Common Notions, we can conclude that sum of the angles $\angle DAC$, $\angle CAB$, and $\angle BAE$ is equal to two right angles.

Since DE is parallel to BC, we know that alternate internal angles are equal (Proposition 9.22) and so $\angle ACB$ equals $\angle DAC$ and $\angle BAE$ equals $\angle ABC$. So by the Common Notions, the interior angles are equal to two right angles. □

9.4.2 Parallelograms

We will now use the theorems regarding parallel lines and the angles of a transverse to prove properties regarding parallelograms and other related quadrilaterals.

Related Content Standards

- (HSG.CO.11) Prove theorems about parallelograms. *Theorems include: opposite sides are congruent, opposite angles are congruent, the diagonals of a parallelogram bisect each other, and conversely, rectangles are parallelograms with congruent diagonals.*

Note that some of the definitions given below are not the same as those given by Euclid, but we will use these definitions for the remainder of the text.

Definition 9.2. A **quadrilateral** is a set of four points A, B, C, D, of which no three are collinear, and four segments AB, BC, CD, DA, of which none intersect at a point other than A, B, C, D.

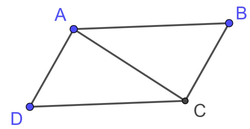
This eliminates the degenerate case for which the sides of the quadrilateral cross.

Definition 9.3. A **parallelogram** is a quadrilateral where the opposite sides are parallel.

Theorem 9.2. *In a parallelogram, the opposite sides are of equal length.*

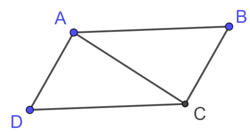
Proof. Let $ABCD$ be a parallelogram. Then AC is a transversal and by Proposition 9.22 we have that $\angle DAC$ equals $\angle BCA$ and $\angle BAC$ equals $\angle DCA$.

Since AC equals CA, Proposition 9.15 implies that $\triangle ABC$ is congruent to $\triangle CDA$ and so AB equals CD and AD equals CB, so the opposite sides are of equal length. □



Theorem 9.3. *If the opposite sides in a quadrilateral are the same length, then the figure is a parallelogram.*

Proof. Let $ABCD$ be a quadrilateral whose opposite sides are the same length. Then we have that AB equals CD, AD equals CB, and AC equals CA.



So by Proposition 9.8 we have that $\triangle ABC$ is congruent to $\triangle CDA$. This implies that angles $\angle DAC$ and $\angle BCA$ are equal and angles $\angle BAC$ and $\angle DCA$ are equal. So by Proposition 9.20, the opposite sides of the quadrilateral are parallel. □

Theorem 9.4. *A quadrilateral is a parallelogram if and only if the diagonals bisect each other.*

Proof. Let $ABCD$ be a convex quadrilateral and let E be the intersection point of the two diagonals. If $ABCD$ were not convex, then it could not be a parallelogram and the diagonals would not intersect. Since this theorem is an if and only if statement, it involves proving both implications.

Let us assume that $ABCD$ is a parallelogram. Since AB is parallel to CD, we know that the angles $\angle EAB$ and $\angle ECD$ are equal and the angles $\angle EBA$ and $\angle EDC$ are equal. Theorem 9.3 implies that AB is equal to DC. So by Proposition 9.15 we know that $\triangle ABE$ is congruent to $\triangle DCE$. This means that AE and CE are equal and BE and DE are equal. So the diagonals bisect each other.

Let us assume that the diagonals bisect each other. We know that vertical angles are equal by Proposition 9.18. Combining this with AE equaling CE and BE equaling DE we use Proposition 9.15 to see that $\triangle ABE$ is congruent to $\triangle CDE$ and $\triangle ADE$ is congruent to $\triangle CBE$. This means that the opposite sides of the quadrilateral are equal and so, by our previous theorem, the quadrilateral is a parallelogram. □

We will now see how parallelograms and rectangles are related. But first we need to choose a definition for a rectangle.

Definition 9.4. A **rectangle** is a quadrilateral where all four angles are the same size.

Notice that we have not defined the rectangle to have angles that are right angles, following the preferential criteria for definitions of minimality and elegance. We will show in the next theorem that the four angles of the same size in a quadrilateral implies that all of the angles are right angles.

Theorem 9.5. *A rectangle has four right angles.*

Proof. Let $ABCD$ be a rectangle. Since all of the angles are the same size, the quadrilateral must be convex. Otherwise, one of the angles would be larger than a right angle and the others less than a right angle. If we let AC be a diagonal, we see that we have triangles $\triangle ABC$ and $\triangle ADC$. The sum of the angles of these triangle are each equal to two right angles by Theorem 9.1. Using the Common Notions, we see that the sum of the angles of the two triangles is equal to the sum of the angles of the quadrilateral. This means that the sum of the angles of the convex quadrilateral must equal four right angles. Since the four angles of the rectangle are equal, each must be a right angle. □

Theorem 9.6. *A rectangle is a parallelogram.*

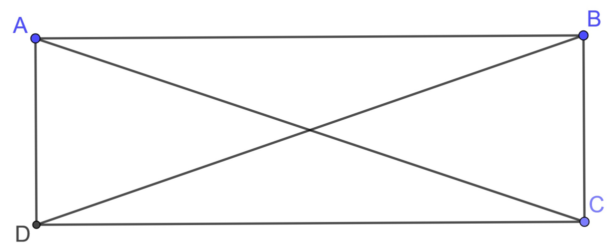
Proof. Let $ABCD$ be a rectangle. We can extend each of the sides of the quadrilateral to infinite lines and see that each of the intersections of lines form right angles. So all of the opposite interior angles are right angles and are therefore equal. Therefore, the opposite sides of the rectangle are parallel and so the rectangle is a parallelogram. □

Theorem 9.7. *A parallelogram is a rectangle if and only if one of the angles is a right angles.*

Proof. The proof will be left as an exercise. □

Theorem 9.8. *A parallelogram is a rectangle if and only if the diagonals are the same length.*

Proof. Let $ABCD$ be a parallelogram with diagonals AC and BD. Since $ABCD$ is a parallelogram, the opposite sides AD and BC are equal.



Since CD equals DC, we know that $\triangle ACD$ is congruent to $\triangle BDC$ if and only if the angles $\angle ADC$ and $\angle BCD$ are equal by Proposition 9.4. We also know that $\triangle ACD$ is congruent to $\triangle BDC$ if and only if the sides AC and BD are equal. Therefore, the sides AC and BD are equal if and only if the angles $\angle ADC$ and $\angle BCD$ are equal.

Since the angles are interior angles, their sum is equal to two right angles. So the diagonals are equal if and only if the angles $\angle ADC$ and $\angle BCD$ are right angles. By Theorem 9.7 we see that the diagonals are equal if and only if the parallelogram is a rectangle. □

Another quadrilateral of interest when discussing parallelograms are rhombi.

Definition 9.5. A **rhombus** is a quadrilateral where all four sides have the same length.

Theorem 9.9. *A rhombus is a parallelogram.*

Proof. This follows directly from Theorem 9.3. □

Theorem 9.10. *A quadrilateral is a rhombus if and only if the diagonals are perpendicular bisectors of each other.*

Proof. The proof will be left as an exercise. □

9.4.3 Exercises

1. Write out the details of the proof of Proposition 9.17.
2. For each of the propositions, use a dynamic geometry app (e.g. GeoGebra or Desmos) to sketch the diagrams and follow the arguments.
3. Prove Theorem 9.7.
4. Prove the following Theorem.

Theorem 9.11. *If one pair of opposite sides in a four sided figure are both equal and parallel, then the figure is a parallelogram.*

5. Prove Theorem 9.10.

6. Prove the following Theorem.

Theorem 9.12. *A quadrilateral is a rhombus if and only if the diagonals bisect all the vertex angles.*

7. Prove the following Theorem.

Theorem 9.13. *If one of the diagonals in a parallelogram bisects one of the vertex angles, then the parallelogram is a rhombus.*

9.5 Similarity of Triangles

Definition 9.6. Two triangles $\triangle ABC$ and $\triangle A'B'C'$ are **similar**, also called equiangular, if they have congruent angles.

$\angle ABC$ equals $\angle A'B'C'$, $\angle BAC$ equals $\angle B'A'C'$, and $\angle BCA$ equals $\angle B'C'A'$.

When two triangles are similar, we will write $\triangle ABC \sim \triangle A'B'C'$.

Theorem 9.14. *Triangles $\triangle ABC$ and $\triangle A'B'C'$ are similar if and only if any two of the corresponding angles are equal.*

Theorem 9.15 (SSS Similarity)**.** *Triangles $\triangle ABC$ and $\triangle A'B'C'$ are similar if and only if the corresponding sides have lengths in the same ratio:*

$$\frac{AB}{A'B'} = \frac{BC}{B'C'} = \frac{AC}{A'C'}$$

Theorem 9.16 (SAS Similarity)**.** *Triangles $\triangle ABC$ and $\triangle A'B'C'$ are similar if and only if two sides have lengths in the same ratio and the angles included between these sides have the same measure.*

9.5.1 Exercises

1.

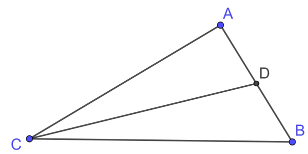
9.6 Centers of Triangles

When we think about a "center" of a triangle, there are many different properties that we may want. We will now explore how some of these properties generate various "centers" of a triangle.

9.6.1 Centroid

The first "center" that we will consider is the point that is the arithmetic mean of all of the points contained inside of the triangle. From a physics perspective, this point is the center of mass, or the balance point, for the triangle. To generate this point, we need to first create some definitions.

Definition 9.7. The **median** of a side of a triangle is the segment connecting the midpoint of the side to the opposite vertex.



So for $\triangle ABC$, if D is the midpoint of AB, then CD is the median of AB.

Theorem 9.17. *Medians of a triangle are concurrent at a point.*

We will give an outline of the proof in a series of lemmas with some key details left to be filled in. We recommend that as you read through the proof you should use a dynamic geometry application to construct the diagrams and to study properties of the constructions.

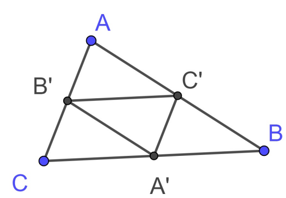
Lemma 9.1. *Let $\triangle ABC$ be given. Let A' be the midpoint of BC, B' be the midpoint of AC, and C' be the midpoint of AB.*

Then $\triangle A'B'C \sim \triangle BAC$ with

$$\frac{A'B'}{BA} = \frac{A'C}{BC} = \frac{B'C}{AC} = 2$$

and triangles $\triangle AB'C'$, $\triangle B'A'C$, $\triangle C'BA'$, and $\triangle A'B'C'$ are all congruent and so are also all similar to $\triangle ABC$ with ratio of $2:1$.

Proof. Let $\triangle ABC$ be given. Let A' be the midpoint of BC, B' be the midpoint of AC, and C' be the midpoint of AB.



Since A' is the midpoint of BC, B' is the midpoint of AC, and angles $\angle BCA$ and $\angle A'CB'$ are equal, by Theorem 9.16 we know that $\triangle A'B'C \sim \triangle BAC$ with

$$\frac{A'B'}{BA} = \frac{A'C}{BC} = \frac{B'C}{AC} = 2.$$

Using similar arguments, we have that $\triangle AB'C' \sim \triangle ACB$ and $\triangle BC'A' \sim \triangle BAC$ with a ratio of $2:1$.

We now know that

$$\frac{A'B'}{AB} = \frac{A'C'}{AC} = \frac{B'C'}{BC} = 2$$

and so by Theorem 9.15 $\triangle A'B'C' \sim \triangle ABC$ with a ratio of $2:1$. □

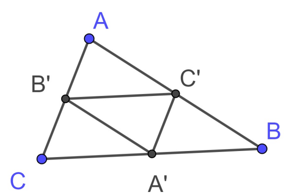
So the segments connecting the midpoints of the sides of a triangle create four congruent triangles that are similar to the original triangle and whose side lengths are half the length of the original.

Related Content Standards

- (HSG.CO.10) Prove theorems about triangles. *Theorems include: measures of interior angles of a triangle sum to* $180°$*; base angles of isosceles triangles are congruent; the segment joining midpoints of two sides of a triangle is parallel to the third side and half the length; the medians of a triangle meet at a point.*

Lemma 9.2. *The segments joining the midpoints of two sides of a triangle is parallel to the third side and half the length.*

Proof. The proof the previous lemma included that the segments joining the midpoints of two sides of a triangle is half the length of the third side. So we only need to prove that it is parallel to the third side.



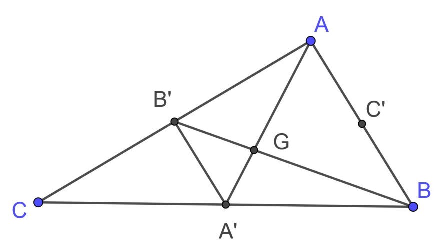
Since the four triangles inside of the original triangle are congruent to one another we know that the angles $\angle BC'A'$ and $\angle B'A'C'$ are equal. Since these are alternate interior angles of the transversal $A'C'$ crossing AB and $A'B'$ we see that $A'B'$ is parallel to AB. We use the same argument to show that the other two segments are parallel to the corresponding sides of the triangle. $\square$

Lemma 9.3. *Let $\triangle ABC$ be given. Let A' be the midpoint of BC, B' be the midpoint of AC, and C' be the midpoint of AB and let G be the point of intersection of AA' and BB'.*

Then $\triangle AGB \sim \triangle A'GB'$ and

$$\frac{AG}{A'G} = \frac{BG}{B'G} = \frac{AB}{A'B'} = 2.$$

Proof. Let $\triangle ABC$ be given. Let A' be the midpoint of BC, B' be the midpoint of AC, and C' be the midpoint of AB and let G be the point of intersection of AA' and BB'.



Since AB is parallel to $A'B'$ we know that $\angle GA'B' = \angle GAB$ and $\angle GB'A' = \angle GBA$ since they are alternate interior angles. Also, $\angle A'GB' = \angle AGB$ because they are vertical angles. So $\triangle AGB \sim \triangle A'GB'$ by the definition of similarity. Since

$$\frac{AB}{A'B'} = 2$$

we have the remainder of the ratios by Theorem 9.15. $\square$

We can use the same arguments to prove the following lemma about the intersection point of AA' and CC'.

Lemma 9.4. *Let $\triangle ABC$ be given. Let A' be the midpoint of BC, B' be the midpoint of AC, and C' be the midpoint of AB and let G' be the point of intersection of AA' and CC'.*

Then $\triangle AG'C \sim \triangle A'G'C'$ and

$$\frac{AG'}{A'G'} = \frac{CG'}{C'G'} = \frac{AC}{A'C'} = 2.$$

Question: Why do we give the point of intersection of AA' and CC' separate names rather than just calling it G?

Since G and G' are both on AA' and $\frac{AG}{A'G} = \frac{AG'}{A'G'} = 2$ we see that $G = G'$.

Question: Why is proving that G and G' are at the same point along a median segment sufficient for proving that $G = G'$? Be as specific as possible.

So the medians AA', BB', and CC' are concurrent.

Definition 9.8. The **centroid** of a triangle is the point of concurrence of the medians of the sides of the triangle.

9.6.2 Circumcenter

The next "center" we consider is the point whose distance is the same from the three vertices of the triangle and so is the center of the circle that circumscribes the triangle. A unique property of this point is that it is not always inside of the triangle. One can prove that the circumcenter is inside of the triangle if and only if the triangle is acute.

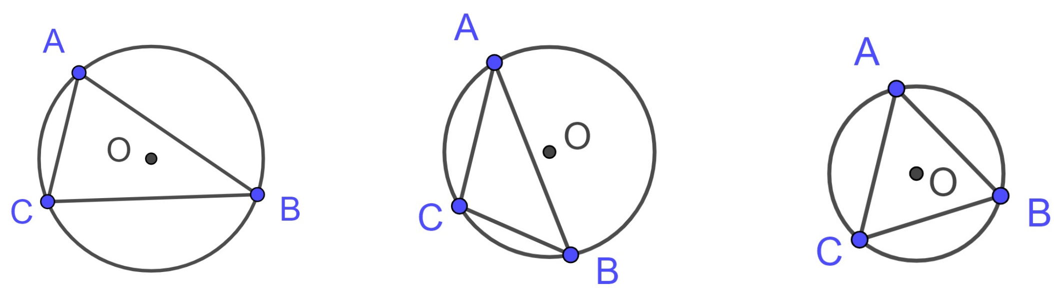


Figure 9.5: Sample Venn diagrams

Theorem 9.18. *The perpendicular bisectors of the three edges of a triangle are concurrent at a point.*

Proof. Let $\triangle ABC$ be given and let O be the point of intersection of the perpendicular bisectors of AB and AC. Because O is on the perpendicular bisector of AB, $AO = BO$. Since O is on the perpendicular bisector of AC, $AO = CO$. By transitivity, $BO = CO$ and so O is on the perpendicular bisector of BC. So the perpendicular bisectors are concurrent. □

Definition 9.9. The **circumcenter** of a triangle is the point of concurrence of the perpendicular bisectors of the sides of the triangle.

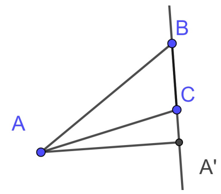
9.6.3 Orthocenter

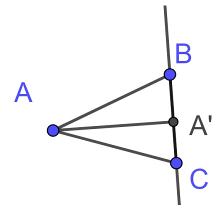
A third point related to the triangle involves the point of intersection of the altitudes of the triangle.

Definition 9.10. The **altitude** of a vertex of a triangle is the segment connecting the vertex of a triangle to a point on the extension of the opposite side of the triangle so that the segment is perpendicular to the opposite side.

In each of the two triangles below, AA' is the altitude of A. In the first triangle, the altitude is outside of the triangle, while it is inside of the triangle in the second example.

In order to prove that the altitudes are concurrent we will first prove a useful theorem regarding concurrency proved by Giovanni Ceva (1648-1734).

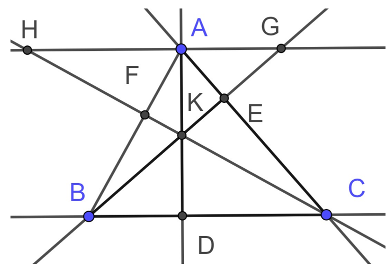




Theorem 9.19 (Ceva's Theorem)**.** *In a triangle $\triangle ABC$, three lines AD, BE, and CF intersect at a single point K if and only if*

$$\frac{AF}{FB} \cdot \frac{BD}{DC} \cdot \frac{CE}{EA} = 1.$$

Proof. We first assume that the three lines AD, BE, and CF are concurrent. We can then extend the lines BE and CF beyond the triangle until they meet GH, the line through A parallel to BC.



By alternate interior angles, $\angle AHF$ and $\angle BCF$ is equal and by vertical angles, $\angle AFH$ and $\angle CFB$ are equal. So by Theorem 9.14, $\triangle AHF \sim \triangle BCF$. Using the same type of arguments, $\triangle AGE \sim \triangle CBE$.

We can also use similar arguments to show that $\triangle AGK \sim \triangle BDK$ and $\triangle CDK \sim \triangle AHK$.

These similar triangles imply the following proportions:

$$\frac{AF}{FB} = \frac{AH}{BC}, \quad \frac{CE}{EA} = \frac{BC}{AG}, \quad \frac{AG}{BD} = \frac{AK}{DK}, \quad \frac{AH}{DC} = \frac{AK}{DK}$$

From the last two proportions we conclude that $\frac{AG}{BD} = \frac{AH}{DC}$ and so $\frac{BD}{DC} = \frac{AG}{AH}$.

Using these proportions we have that

$$\frac{AF}{FB} \cdot \frac{BD}{DC} \cdot \frac{CE}{EA} = \frac{AH}{BC} \cdot \frac{AG}{AH} \cdot \frac{BC}{AG} = 1.$$

In order to prove the converse, we will assume

$$\frac{AF}{FB} \cdot \frac{BD}{DC} \cdot \frac{CE}{EA} = 1$$

and prove that this implies AD, BE, and CF are concurrent.

Assume that K is the point of intersection of BE and CF and draw the line AK until its intersection with BC at a point D'. Then, from the just proven part of the theorem it follows that

$$\frac{AF}{FB} \cdot \frac{BD'}{D'C} \cdot \frac{CE}{EA} = 1.$$

On the other hand, it's given that

$$\frac{AF}{FB} \cdot \frac{BD}{DC} \cdot \frac{CE}{EA} = 1,$$

so that $\frac{BD'}{D'C} = \frac{BD}{DC}$ implying that D and D' are the same point. So the segments are concurrent. □

Theorem 9.20. *The lines extending altitudes of the vertices of a triangle are concurrent.*

Proof. Let $\triangle ABC$ be given. Let A' be point on the line BC so that AA' is the altitude of A, B' be the point on the line AC so that BB' is the altitude of B, and C' be the point on the line AB so that CC' is the altitude of C.

We have that triangles $\triangle ACA'$ and $\triangle BCB'$ are similar because they both have a right angle and share an angle at C. So $\frac{CB'}{A'C} = \frac{BB'}{AA'}$. Similarly,

$$\frac{AC'}{B'A} = \frac{CC'}{BB'} \quad \text{and} \quad \frac{BA'}{C'B} = \frac{AA'}{CC'}.$$

Therefore,

$$\frac{AC'}{C'B} \cdot \frac{BA'}{A'C} \cdot \frac{CB'}{B'A} = 1$$

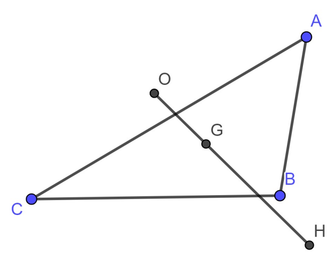
and Ceva's Theorem implies that the lines are concurrent. □

Definition 9.11. The point of concurrency of the extensions of the altitudes of the vertices of a triangle is called the **orthocenter**.

9.6.4 Euler Line

Leonard Euler showed in 1765 that the centroid, circumcenter, and orthocenter of a triangle are colinear and the line through these three points is called the Euler line of the triangle.

Theorem 9.21 (Euler Line)**.** *Given a triangle $\triangle ABC$ with centroid G, circumcenter O, and orthocenter H, the points G, H, and O are colinear with G between O and H and is twice as far from the orthocenter as from the circumcenter, $GH = 2GO$.*



9.6.5 Exercises

1. Reflection questions about the centroid.

   a. At what points in the proof, if any, were axioms or common definitions used to illuminate something?
   b. At what points in the proof, if any, were definitions used to illuminate something?
   c. When, if at all, were propositions or theorems used to illuminate something?
   d. How were theorems used in the proof? (i.e., focus on what work went into using the theorem).
   e. In geometry, students often struggle with the fact that a picture of the thing they are proving "shows" the result, so it seems pointless to write everything out. Why do you think we insist on doing it this way anyway?

2. Prove that the angle bisectors of a triangle are concurrent. This point of concurrency is called the **incenter** of the triangle.

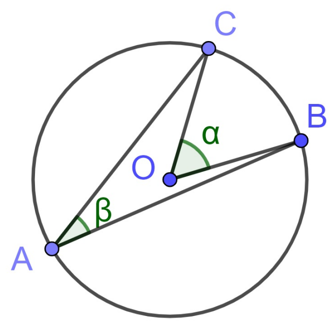
9.7 Circle Theorems

In Book IV of Euclid's Elements [Heath, 1908b] Euclid returns to constructions involving regular polygons and circles. Before we look at these constructions we need to first define some terms and prove some basic properties of circles.

Definition 9.12. An **inscribed angle** is an angle formed by two chords in a circle which have a common endpoint. This common endpoint forms the vertex of the inscribed angle. The other two endpoints define what we call an intercepted arc on the circle.

Definition 9.13. A **central angle** is any angle whose vertex is located at the center of a circle. A central angle necessarily passes through two points on the circle, which in turn divide the circle into two arcs: a major arc and a minor arc. The minor arc is the smaller of the two arcs, while the major arc is the bigger. We define the arc angle to be the measure of the central angle which intercepts it.

The inscribed angle and central angle related to the same arc are related in the following way.



Theorem 9.22. *In a circle, the measure of an inscribed angle is half the measure of the central angle with the same intercepted arc.*

Proof. Let O be the center of a circle with unknown radius. Let A be the common end point of chords AB and AC on the circle and let α be the measure of $\angle COB$ and β be the measure of $\angle CAB$.

OB and OC are the same length because they are both radii of the same circle. Therefore, $\triangle BOC$ is isosceles and $\angle OCB$ and $\angle OBC$ are equal and both equal $\frac{180^\circ - \alpha}{2}$.

We also see that $\triangle OAC$ is isosceles and we can let x be the measure of angles $\angle OAC$ and $\angle OCA$. Similarly $\triangle OBA$ is isosceles and we can let y be the measure of the angles $\angle OBA$ and $\angle OAB$.

Then $\beta = x + y$, the measure of $\angle ACB$ is $x + \frac{180^\circ - \alpha}{2}$ and the measure of $\angle ABC$ is $y + \frac{180^\circ - \alpha}{2}$. Since the sum of the angles of a triangle equal two right angles,

$$(x+y) + \left(x + \frac{180^\circ - \alpha}{2}\right) + \left(y + \frac{180^\circ - \alpha}{2}\right) = 180^\circ.$$

Reordering the terms leads to $2(x + y) = \alpha$, and since $\beta = x + y$, we have our result. □

If the central angle is a straight line forming a diameter to the circle, the central angle equals two right angles and we have the following.

Theorem 9.23. *Any angle inscribed in a semi-circle is a right angle.*

We are now ready to return to the constructions of Euclid, which we will leave as exercises.

Related Content Standards

- (HSG.CO.13) Construct an equilateral triangle, a square, and a regular hexagon inscribed in a circle.

9.7.1 Exercises

1. Construct the inscribed circle of a triangle, and prove that it is such.
2. Construct the circumscribed circle of a triangle, and prove that it is such.
3. Construct a tangent line from a point outside a given circle to the circle, and prove that it is such.
4. Construct an equilateral triangle inscribed in a circle, and prove that it is such.
5. Construct a square inscribed in a circle, and prove that it is such.
6. Construct a regular hexagon inscribed in a circle, and prove that it is such.

Chapter 10

Measurement

Measurement is an overlap between the areas of mathematics, science, and commerce designed to help us quantify how much of something we have. We commonly do this with inches, meters, square yards, liters, gallons, etc. The introduction of these topics to children takes place primarily in elementary school. However, middle and high school students often continue to struggle with measurement relationships and mathematical modeling related to high school geometry relies heavily on students' knowledge of measurement. Thus, we quickly review the fundamental notions of measurement and explore how these ideas are used to build the common formulas we use in middle and high school problem solving.

10.1 Units

The first things children measure are usually the number of discrete objects, such as a number of toys. This concept of measuring objects is then extended to distances, areas, volumes, and masses throughout the elementary curriculum. In each of these instances the concept of a unit of measurement is the essential topic.

10.1.1 Units of Length

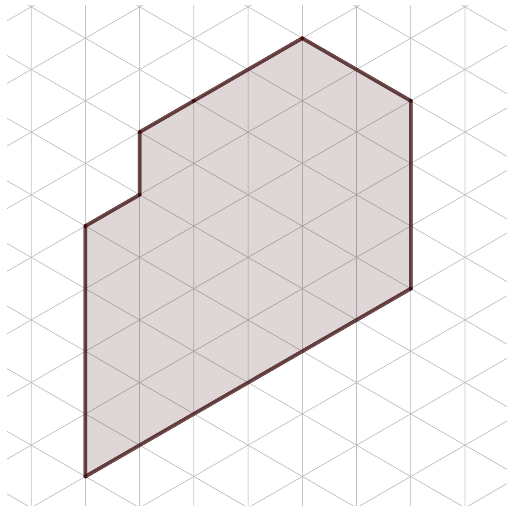
In first and second grade, students are introduced to the measurement of the length of certain objects. Initially this is done using shorter objects and then counting the number of the shorter objects contained in the larger object. This parallels the units of measurement based on the length of various parts of the body such as the forearm (cubits), a part of a finger (inches), or the length of the foot (feet). Students are then led to recognize the need for standard units of measurement in order to have a common vocabulary. Historically this standardization process was based on the measurements related to the king. However, when a monarch changed, so did the units of measurement. This standardization process led to the creation of the metric system in the eighteenth century and the current International System of Units (SI) where units of length are derived from a meter.

Related Content Standards

- (1.MD.2) Express the length of an object as a whole number of length units, by laying multiple copies of a shorter object (the length unit) end to end; understand that the length measurement of an object is the number of same-size length units that span it with no gaps or overlaps. Limit to contexts where the object being measured is spanned by a whole number of length units with no gaps or overlaps.
- (2.MD.1) Measure the length of an object by selecting and using appropriate tools such as rulers, yardsticks, meter sticks, and measuring tapes.

10.1.2 Units of Area

In third grade, students are introduced to the idea of measuring two-dimensional figures by determining the number of a certain two-dimensional objects that can fit inside the object being measured. Using isometric graph paper (made with equilateral triangles) one can determine the number of triangles that fit inside of an object. So in the figure below, the polygon in the figure below has an area of 54 triangles.

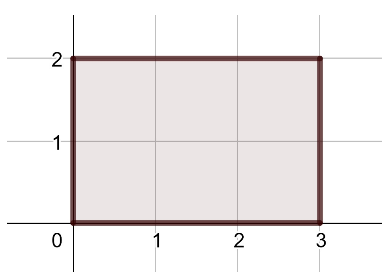


As a way to simplify the process of measuring area and to make a connection to units of length, the standard unit of measure for area is a square whose sides are 1 unit of the corresponding length measurement. So if one is measuring a planar figure whose edges would be measured using inches, the area is measured using squares whose side lengths are 1 inch. These units are then called square inches (or in^2).

Related Content Standards

- (3.MD.5) Recognize area as an attribute of plane figures and understand concepts of area measurement.
  a) A square with side length 1 unit, called "a unit square," is said to have "one square unit" of area, and can be used to measure area.
  b) A plane figure which can be covered without gaps or overlaps by n unit squares is said to have an area of n square units.
- (3.MD.6) Measure areas by counting unit squares (square cm, square m, square in, square ft, and improvised units).

By using square units as the system of measurement, one can compute the area of a rectangle by taking the length measurements of adjacent sides and finding the product. So that a rectangle with sides of 2 cm and 3 cm would have an area of 6 square cm.



By focusing on the unit of measure corresponding to the actual square it helps students to differentiate between measurements related to perimeter and area, a common source of challenges for students.

10.1.3 Units of Volume

In a similar way to area, volume is often measured by the number of unit cubes that can fit inside of a three dimensional object. When volume is measured in terms of these cubes, the volume of a rectangular prism can be found by multiplying the edge lengths of the prism.

Related Content Standards

- (5.MD.3) Recognize volume as an attribute of solid figures and understand concepts of volume measurement.
  a. A cube with side length 1 unit, called a "unit cube," is said to have "one cubic unit" of volume, and can be used to measure volume.
  b. A solid figure which can be packed without gaps or overlaps using n unit cubes is said to have a volume of n cubic units.
- (5.MD.4) Measure volumes by counting unit cubes, using cubic cm, cubic in, cubic ft, and improvised units.
- (5.MD.5) Relate volume to the operations of multiplication and addition and solve real world and mathematical problems involving volume.
  a. Find the volume of a right rectangular prism with whole-number side lengths by packing it with unit cubes, and show that the volume is the same as would be found by multiplying the edge lengths, equivalently by multiplying the height by the area of the base. Represent threefold whole-number products as volumes, e.g., to represent the associative property of multiplication.
  b. Apply the formulas $V = l \times w \times h$ and $V = b \times h$ for rectangular prisms to find volumes of right rectangular prisms with whole-number edge lengths in the context of solving real world and mathematical problems.
  c. Recognize volume as additive. Find volumes of solid figures composed of two non-overlapping right rectangular prisms by adding the volumes of the non-overlapping parts, applying this technique to solve real world problems.
- (6.G.2) Find the volume of a right rectangular prism with fractional edge lengths by packing it with unit cubes of the appropriate unit fraction edge lengths, and show that the volume is the same as would be found by multiplying the edge lengths of the prism. Apply the formulas $V = l \times w \times h$ and $V = b \times h$ to find volumes of right rectangular prisms with fractional edge lengths in the context of solving real-world and mathematical problems.

Another standard unit of measure for volume involves liquid measures such as liters, cups, or gallons. Such units make more sense in terms of the units being directly tied to volume, but are not as easy to connect to the linear or area units. In order to improve this connection the international standard units are established in a way to make this connection, as the cubic centimeter is equivalent to a milliliter.

Related Content Standards

- (3.MD.2) Measure and estimate liquid volumes and masses of objects using standard units of grams (g), kilograms (kg), and liters (l). Add, subtract, multiply, or divide to solve one-step word problems involving masses or volumes that are given in the same units, e.g., by using drawings (such as a beaker with a measurement scale) to represent the problem.

10.1.4 Units of Angle Measurements

Another type of unit needed in geometric mathematical modeling is a way to measure the size of an angle. One basic premise in creating a unit to measure an angle is to divide a circle into smaller pieces, with the primary unit being a degree. The determination to have 360 degrees in a circle may be based on an idea of 360 days in a year (with rounding). Another possibility for the origin is that a circle can be divided into six equilateral triangles. Each of these triangles is then divided into 60 parts, based on the base 60 number system of the Babylonians.

A second idea is to base the measurement of an angle on the length on a circle swept out by the angle. Using this foundation we can define a radian to be the angle swept out of a circle so that the arclength on the circle is the same as the radius of a circle. Since the radian is a ratio of lengths, it is a method of measuring that is dimensionless and so we do not write the unit name most of the time.

Related Content Standards

- (4.MD.5) Recognize angles as geometric shapes that are formed wherever two rays share a common endpoint, and understand concepts of angle measurement:
  a. An angle is measured with reference to a circle with its center at the common endpoint of the rays, by considering the fraction of the circular arc between the points where the two rays intersect the circle. An angle that turns through 1/360 of a circle is called a "one-degree angle," and can be used to measure angles.
  b. An angle that turns through n one-degree angles is said to have an angle measure of n degrees.
- (HSF.TF.1) Understand radian measure of an angle as the length of the arc on the unit circle subtended by the angle.

10.1.5 Unit Conversions and Dimensional Analysis

In many modeling circumstances information is given in one type of unit and needs to be converted to another type of unit. Since many units share the same value of zero, there is a linear relationship between the units. For example, we know that 3 feet is the same as 1 yard. So the expression $\frac{3 \text{ feet}}{1 \text{ yard}}$ is equivalent to the number 1. Then if we want to convert 25 feet into yards we see that

$$25 \text{ feet} = 25 \text{ feet} \cdot 1 = \frac{25 \text{ feet}}{1} \cdot \frac{1 \text{ yard}}{3 \text{ feet}} = \frac{25}{3} \text{ yards} = 8\frac{1}{3} \text{ yards}.$$

Using these concepts we can use conversions between linear units to find conversions between area and volume units. So 1 square yard in terms of square feet,

$$1 \text{ yd}^2 = \frac{1 \text{ yd}^2}{1} \cdot \frac{3 \text{ ft}}{1 \text{ yd}} \cdot \frac{3 \text{ ft}}{1 \text{ yd}} = 9 \text{ ft}^2,$$

we see that a square yard is equivalent to 9 square feet. Similarly, we can see that 1 cubic yard is equivalent to 27 cubic feet.

Related Content Standards

- (5.MD.1) Convert among different-sized standard measurement units within a given measurement system (e.g., convert 5 cm to 0.05 m), and use these conversions in solving multi-step, real world problems.
- (7.G.1) Solve problems involving scale drawings of geometric figures, including computing actual lengths and areas from a scale drawing and reproducing a scale drawing at a different scale.

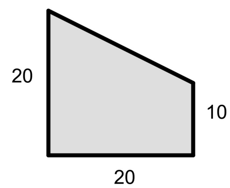
These methods of converting units can also be applied to rates of change such as speed where we can find the equivalent of 55 miles per hour in meters per second,

$$\frac{55 \text{ miles}}{1 \text{ hour}} = \frac{55 \text{ miles}}{1 \text{ hour}} \cdot \frac{5280 \text{ feet}}{1 \text{ mile}} \cdot \frac{1 \text{ meter}}{3.28 \text{ feet}} \cdot \frac{1 \text{ hour}}{60 \text{ minutes}} \cdot \frac{1 \text{ minute}}{60 \text{ seconds}} = 24.59 \frac{\text{meters}}{\text{second}}.$$

10.1.6 Exercises

1. Adrian is planting seeds 1 foot apart in a grid pattern in the horizontal and vertical directions. If he plants 4 seeds in the horizontal direction and 7 seeds in the vertical direction, what is the area of the plot he is working with? (Assume he used the maximum space available.)

2. Samantha is planting a vinyard of grape vines. She knows that each vine needs to have a rectangular piece of land that is 1 foot by 3 feet in a unit that we will call a vine-block. What is the area of the field below, with linear measurements given in feet, in terms of vine-blocks? Why is the unit not 'squared'?



1. If a circle is measured to be 3 in^2, do we need the squares to be visible to know what that means? Explain your reasoning to a middle school student.

2. For each of the following exercises pay close attention to units and precision of language.

   a. You are buying carpet to cover a room that measures 16 ft by 28 ft. The carpet cost $22 per square yard. How much will the carpet cost?
   b. A cargo container has the following internal dimensions
      - L: 39 feet, $\frac{3}{8}$ inches
      - W: 7 feet, $8\frac{1}{8}$ inches
      - H: 7 feet, $9\frac{5}{8}$ inches

      Find the volume inside of the container ship in cubic feet, cubic yards, and cubic meters.
   c. A car travels 30 km in 15 minutes. How fast is it going in kilometres per hour? In metres per second?

3. How does the relationship between linear and area units in this section help answer the question "when I multiply $\frac{1}{3}$ times $\frac{1}{3}$ I get $\frac{1}{9}$ following the rule, but that cannot be because when I multiply numbers together, they are supposed to get bigger."

4. Find some graph paper. Draw polygons with an area equal to 6 unit squares such that all the perimeter lines of your polygons lie only on the grid lines of the graph paper.

   a. How many unique polygons (the polygon is not another one under rotation or reflection) can you construct?
   b. Find the perimeter of each of your polygons. What do these tell you, if anything, about the relationship between area and perimeter?
   c. If the rules were changed so that you only had to keep the area as six square units, what is the range of possible perimeters for your polygon?

10.2 Decomposing and Composing

A key mathematical way of thinking is the ability to work with mathematical structures to decompose and recompose mathematical objects in multiple ways in order to better understand properties of the objects and become proficient in related mathematical computation. In early elementary grades, students decompose natural numbers into different sums of numbers in order to improve computational proficiency. This is particularly important in student understanding and use of the base 10 number system when asked to rewrite expressions like 14×32 as

$$(10 + 4) \times (30 + 2) = (10 \times 30) + (4 \times 30) + (10 \times 2) + (4 \times 2).$$

Similarly we are able to better understand geometric shapes by decomposing the shape into shapes that we already understand, such as rectangles or rectangular prisms. We then use the additive properties of the units of measurement to better understand these new objects.

Related Content Standards

- (3.MD.7) Relate area to the operations of multiplication and addition.

  d. Recognize area as additive. Find areas of rectilinear figures by decomposing them into non-overlapping rectangles and adding the areas of the non-overlapping parts, applying this technique to solve real world problems.

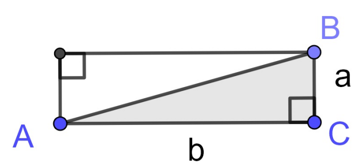
- (4.MD.7) Recognize angle measure as additive. When an angle is decomposed into non-overlapping parts, the angle measure of the whole is the sum of the angle measures of the parts. Solve addition and subtraction problems to find unknown angles on a diagram in real world and mathematical problems, e.g., by using an equation with a symbol for the unknown angle measure

- (5.MD.5) Relate volume to the operations of multiplication and addition and solve real world and mathematical problems involving volume.

  c. Recognize volume as additive. Find volumes of solid figures composed of two non-overlapping right rectangular prisms by adding the volumes of the non-overlapping parts, applying this technique to solve real world problems.

10.2.1 Area

In the prior section we saw that the area of a rectangle can be found by multiplying the lengths of the two sides, $A = l \times w$. We can use this to find a similar expression for the area of a right triangle. If we start with a right triangle $\triangle ABC$ we can create a rectangle by joining to it a copy of itself. We then see that the area of a right triangle is equal to half of the product of the lengths of the two legs of right triangle.

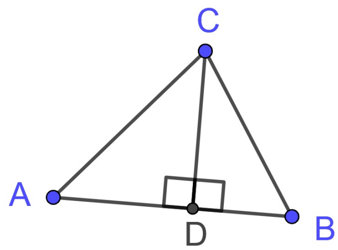


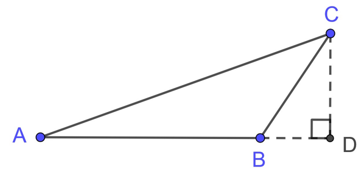
If we have a triangle $\triangle ABC$ that is not a right triangle, we can create an altitude from C to the opposite side of the triangle.

If the altitude is inside of the triangle we see that the original triangle is composed of two right triangles and we can derive that the area of the triangle equals half the product of the length of the altitude and the length the opposite side.

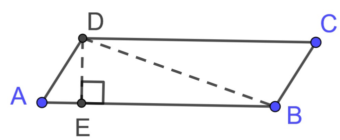
If the altitude is outside of the triangle we can see that the area of the original triangle equals the area of the right triangle $\triangle ADC$ minus the area of the right triangle $\triangle BDC$ giving the area as half the length of the altitude times the length of the opposite side of the triangle.

So for any triangle we have that the area is half the product of the side of the triangle and the height of the triangle from that side, $A = \frac{1}{2}b \cdot h$.





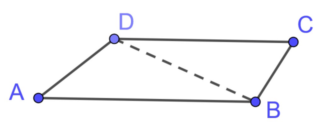
We can now use these areas of rectangles and triangles to find the areas of other two-dimensional objects. If $ABCD$ is a parallelogram, we can divide the quadrilateral into two congruent triangles $\triangle ABD$ and $\triangle CDB$. Each of the triangles has an area of half the product of the length of one set of parallel sides and the length of the distance between those sides. So the area of the parallelogram is given by the product of the length of each set of parallel sides and the distance between the sides.



If $ABCD$ is a trapezoid with sides AB and CD being parallel we can sketch the diagonal BD and see that the area of the quadrilateral is given by

$$\frac{1}{2}b_1 \cdot h + \frac{1}{2}b_2 \cdot h = \frac{1}{2}(b_1 + b_2) \cdot h$$

where b_1 and b_2 are the lengths of the two parallel sides and h is the distance between those sides.



Related Content Standards

- (6.G.1) Find the area of right triangles, other triangles, special quadrilaterals, and polygons by composing into rectangles or decomposing into triangles and other shapes; apply these techniques in the context of solving real-world and mathematical problems.

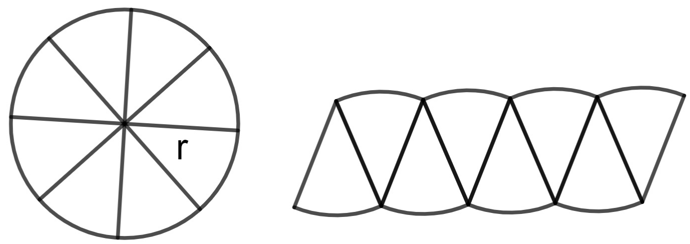
The constant ratio between the circumference and diameter of a circle, labeled as π, allows us to generate formulas for the circumference in terms of both the diameter and radius of a circle, $C = \pi d = 2\pi r$. In order to understand the area of the circle we can divide the circle up into an even number of wedges of equal size. We can then rearrange the wedges as in the figure below.

Then we see that as the number of wedges increase, the figure gets closer to a rectangle whose height equals the radius and the length of one of the other sides is half the circumference of the circle. Using an informal limiting argument, we can then see that the area of the circle is given by

$$A = r \cdot \frac{1}{2}C = r \cdot \frac{1}{2}(2\pi r) = \pi r^2.$$

Related Content Standards

- (7.G.4) Know the formulas for the area and circumference of a circle and use them to solve problems; give an informal derivation of the relationship between the circumference and area of a circle.



10.2.2 Volume

We can follow the same lines of reasoning of decomposing and composing to determine the volumes of three dimensional figures. Archimedes used such ideas in the 3rd century BC with Bonaventure Cavalieri building on these ideas in the mid-1600's.

Theorem 10.1 (Cavalieri's Principle)**.** *Suppose two regions in three-space (solids) are included between two parallel planes. If every plane parallel to these two planes intersects both regions in cross-sections of equal area, then the two regions have equal volumes.*

These ideas are based on slicing the solid into infinitesimal slices and then adding up the volume of these slices. In the special case where the solid is a prism or cylinder each cross-section has an equal area. So Cavlieri's prinicple implies that the volume of prisms or cylinders equals the area of each cross-section times the distance between the two parallel plane, $V = A \cdot h$.

In order to find the volume of a cone or pyramid whose height is d, we let $A(h)$ be the area of the cross-section h units from the base of the solid. Since the solid is a cone or pyramid, the area at each cross-section is proportional to the distance from the base relative to the height,

$$A(h) = \left(1 - \frac{h}{d}\right)^2 \cdot A(0),$$

for $0 \leq h \leq d$.

We can now use the method of exhaustion of the ancient Greeks, as a precursor to calculus, to give an informal argument for the volume of the solid. We know that the solid has a volume less than the volume of the corresponding prism or cylinder, $d \cdot A(0)$. We can get a better estimate for the solid by approximating it by multiple cylinders. If we use 4 cylinders to approximate the volume, we see that the sum of these four volumes is given by

$$\frac{d}{4} \cdot A(0) + \frac{d}{4} \cdot A\left(\frac{d}{4}\right) + \frac{d}{4} \cdot A\left(\frac{2d}{4}\right) + \frac{d}{4} \cdot A\left(\frac{3d}{4}\right).$$

This can be simplified to

$$\frac{d}{4}\left(1 + \left(1 - \frac{1}{4}\right)^2 + \left(1 - \frac{2}{4}\right)^2 + \left(1 - \frac{3}{4}\right)^2\right) \cdot A(0)$$

$$= \frac{d}{4} \cdot \left(\left(\frac{1}{4}\right)^2 + \left(\frac{2}{4}\right)^2 + \left(\frac{3}{4}\right)^2 + 1\right) \cdot A(0).$$

So the area of the solid is approximated by $\frac{30}{64} \cdot d \cdot A(0)$ when using 4 prisms. If we extend this to a larger number of divisions we will use information about the sum of the first n squares,

$$P_n = \frac{n(n+1)(2n+1)}{6}.$$

So if we approximate the solid by n prisms we have the volume as

$$V \approx \frac{1}{n}\left(\sum_{i=1}^{n}\left(\frac{i}{n}\right)^2\right)\cdot d\cdot A(0) = \frac{n(n+1)(2n+1)}{6n^3}\cdot d\cdot A(0).$$

So as the number of prisms increases the volume approximation of the cone or pyramid approaches the true volume,

$$V = \frac{1}{3}\cdot d\cdot A(0).$$

One can then use similar slicing arguments to show that the volume of a sphere is given by

$$V = \frac{4}{3}\pi r^3.$$

Related Content Standards

- (8.G.9) Know the formulas for the volumes of cones, cylinders, and spheres and use them to solve real-world and mathematical problems.
- (HSG.GMD.1) Give an informal argument for the formulas for the circumference of a circle, area of a circle, volume of a cylinder, pyramid, and cone. Use dissection arguments, Cavalieri's principle, and informal limit arguments.
- (HSG.GMD.2) Give an informal argument using Cavalieri's principle for the formulas for the volume of a sphere and other solid figures.
- (HSG.GMD.3) Use volume formulas for cylinders, pyramids, cones, and spheres to solve problems.

10.2.3 Related Horizon Content Knowledge

These concepts of decomposing and recomposing using Cavalieri's principle and limits is the foundation for integration theory. So it is important that any students planning to go into fields of science or engineering get experience with this process when they learn about geometric volumes and areas.

10.2.4 Exercises

1. Use Cavalieri's Principle and prisms to give an informal argument for the volume of a sphere.

2. One common formula for the circumference of a circle is $C = 2\pi r$ and the formula for the area of a circle is $A = \pi r^2$. Students often confuse these two, as they both contain a 2, π, and r. Using the relationship between length and area, what is one way to tell which is which?

3. Perimeter and Areas of polygons.

    a. Create a table with the following headings: Shape name, generic drawing, perimeter formula, area formula. Fill in the table for the following shapes: triangle, square, rectangle, regular n-gon. For the generic drawing, include labels for any variables you use in your perimeter or area formulas.
    b. Reflect on your perimeter formulas. Do you ever multiply two or more dimensions together? Why or why not?
    c. Reflect on your area formulas. Do you ever multiple 3 dimensions together? Why or why not?
    d. Notice that when you multiply dimensions for an area formula the two dimensions are perpendicular to each other. Why does this make sense?

10.3 Measurements of Triangles

A key aspect of measurement in geometry involves the relationships between the lengths of sides of a triangle.

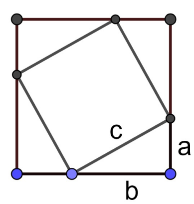
Related Content Standards

- (8.G.6) Explain a proof of the Pythagorean Theorem and its converse.
- (8.G.7) Apply the Pythagorean Theorem to determine unknown side lengths in right triangles in real-world and mathematical problems in two and three dimensions.
- (8.G.8) Apply the Pythagorean Theorem to find the distance between two points in a coordinate system.

Theorem 10.2 (Pythagorean Theorem)**.** *For a right triangle with side lengths a and b, and hypotenuse c,*

$$a^2 + b^2 = c^2.$$

Proof. For any right triangle with side lengths a and b, and hypotenuse c, we construct a square of side lengths $a + b$. We can then construct a figure of four copies of the original right triangle, so that each side length is composed of one copy of each leg of the triangle.



Since we have right triangles on each corner of the square, we can use the properties of congruent triangles to show that the four triangles are congruent, meaning that the interior quadrilateral is a rhombus.

Since the four triangles are congruent, the two angles at each point of intersection between the two quadrilaterals must add to a right angle. So the interior quadrilateral is also a square.

So the area of the larger square can be written in two different ways, as a single square and as the sum of the smaller square and the four triangles,

$$(a + b)^2 = c^2 + 4 \cdot \left(\frac{1}{2}ab\right).$$

Rearranging the terms we have that

$$a^2 + 2ab + b^2 = c^2 + 2ab$$

and so we have that

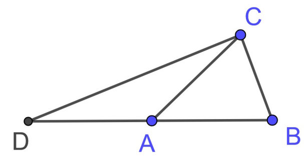
$$a^2 + b^2 = c^2.$$

□

Proposition 10.1 (Euclid's Proposition 20: Triangle Inequality)**.** *In any triangle two sides taken together in any manner are greater than the remaining one.*

Proof. Let $\triangle ABC$ be a triangle.

Extend BA through to the point D so that DA is equal to CA. Create the finite segment DC to create the triangle $\triangle ACD$. Since AD equals AC, $\triangle ACD$ is isosceles and so the angles $\angle ACD$ and $\angle ADC$ are equal. So $\angle BCD$ is greater than $\angle ADC$.



Since $\triangle DCB$ is a triangle having the angle $\angle BCD$ greater than the angle $\angle BDC$, and the side opposite the greater angle is greater, therefore DB is greater than BC.

But DA equals AC, therefore the sum of BA and AC is greater than BC.

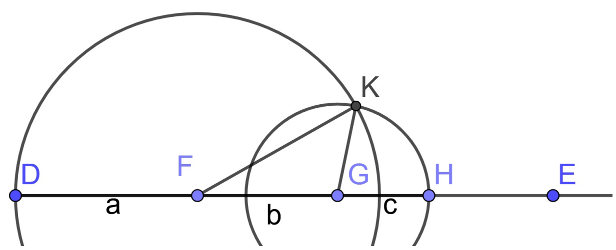
Similarly we can prove that the sum of AB and BC is also greater than CA, and the sum of BC and CA is greater than AB.

Therefore in any triangle the sum of any two sides is greater than the remaining one. □

Proposition 10.2 (Euclid's Proposition 22)**.** *Out of three straight lines, which are equal to three given straight lines, to construct a triangle: thus it is necessary that two of the straight lines taken together in any manner should be greater than the remaining one.*

Proof. Let the three given straight lines be a, b, and c, and let the sum of any two of these be greater than the remaining one, namely, a plus b greater than c, a plus c greater than b, and b plus c greater than b.

Set out a straight line DE, terminated at D but of infinite length in the direction of E. Make DF equal to a, FG equal to b, and GH equal to c.



Construct the circle centered at F through the point D and the circle centered at G through the point H. The two circles intersect due to the relationships between a, b, and c. Let K be one of the points of intersection.

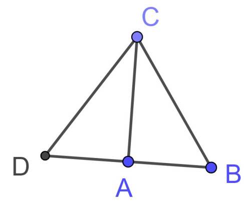
Then the triangle $\triangle FKG$ has the desired side lengths. □

Theorem 10.3 (Converse of Pythagorean Theorem)**.** *For any three positive numbers a, b, and c such that $a^2 + b^2 = c^2$, there exists a triangle with sides a, b and c, and every such triangle has a right angle between the sides of lengths a and b.*

Proof. Let a, b, and c be three positive numbers such that $a^2+b^2 = c^2$. Since $(a+b)^2 = a^2+2ab+b^2 = c^2+2ab$, $c < a + b$. Since $(a - b)^2 = a^2 - 2ab + b^2 = c^2 - 2ab$, $a - b < c$.

By Proposition 10.2 we know that a triangle with sides a, b, and c exists. We will label the triangle $\triangle ABC$ with AB having length a, AC having length b, and BC having length c.

Let D be a point such that AD is perpendicular to AC and AD equals AB. Since triangle $\triangle ACD$ is a right triangle, the Pythagorean theorem implies that CD must have side length c and so is equal to BC. Hence the side lengths of $\triangle ADC$ all equal the side lengths of $\triangle ABC$. So $\triangle ADC$ is congruent to $\triangle ABC$ by Proposition 9.8. Since the angle $\angle DAC$ is a right angle, so is $\angle BAC$. □



10.3.1 Exercises

1. Consider an equilateral triangle with a 1 unit side length. Sketch such a triangle and draw an altitude. Find the exact height of the altitude.
2. Consider a right isosceles triangle with legs 1 unit long. Find the exact length of the hypotenuse.
3. A square is inscribed with a circle such that the four corners of the square sit on the circumference of the circle. The square has side length 1 unit in length. Find the exact diameter of the circle.
4. A rectangle is inscribed within a circle such that the four corners of the rectangle lie on the circle's circumference. The radius of the circles is 1 unit long. If the base of the rectangle is 1.2 units long, find the height of the rectangle.
5. An equilateral triangle is inscribed in a circle of radius r. Using only algebra and Pythagorean relationships (e.g., no trigonometric functions), find an expression for the length of the side length of the triangle in terms of r.

10.4 Distance

We will now move to a generalization of the concept of length using a more rigorous mathematical definition.

Definition 10.1. A **metric**, or **distance**, on a set X is a function

$$d : X \times X \to [0, \infty)$$

such that for all $x, y, z \in X$ the following three axioms are satisfied:

- (D1.) $d(x, y) = 0 \Leftrightarrow x = y$ (identity of indiscernibles)
- (D2.) $d(x,y)=d(y,x) $ (symmetry)
- (D3.) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

If we have equality in the triangle inequality, $d(x, y) = d(x, z) + d(z, y)$, we say that x,y, and z are **colinear** and z is **between** x and y.

If X is our standard plane, $\mathbb{R}^2$, the function $d_2 : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$ defined by

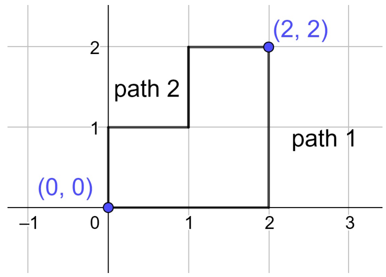
$$d_2\left((x_1, y_1), (x_2, y_2)\right) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

satisfies these axioms and is called the Euclidean distance on the plane because it corresponds with the geometry of Euclid's Elements.

An alternate distance on the plane can be given by $d_1 : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$ defined by

$$d_1\left((x_1, y_1), (x_2, y_2)\right) = |x_2 - x_1| + |y_2 - y_1|$$

and is often called the taxicab distance as it measures the distance traveled along a grid. Hermann Minkowski used this distance in his exploration of non-Euclidean geometries in his study of relativity in the late 1800s.



Unlike the Euclidean distance, the taxicab distance allows multiple paths between two points that have minimal length. We can see below that the distance between the points $(0, 0)$ and $(2, 2)$ is 4, and that both paths given have this minimal length.

However, any point, C, on the line segment connecting two points, A and B, can be considered between A and B since $d(A, C) + d(C, B) = d(A, B)$ for every point on the line segment. This means that the lines in Euclidean geometry and the lines in the taxicab geometry are the same but do not have all of the same properties.

A third distance that has a very natural derivation is $d_\infty : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$ defined by

$$d_\infty \left((x_1, y_1), (x_2, y_2)\right) = \max\left(|x_2 - x_1|, |y_2 - y_1|\right)$$

and is often called the Chebyshev distance after Pafnuty Chebyshev who was the first to systematically study the distance in the mid 1800s.

10.4.1 Circles and Lines

One way that we can see the difference between these three metrics is to look at the circles created by the metrics.

Definition 10.2. Let X be a set with a distance d. A circle centered at a point $c \in X$ with radius r is

$$C_{(c,r)} = \{x \in X \mid d(x, c) = r\}.$$

Since a circle is defined as the set of points equidistant from a fixed point, we see that circles are defined by the distance. Hence when we look at the circles that correspond to the three distances described above we have different sets of points.

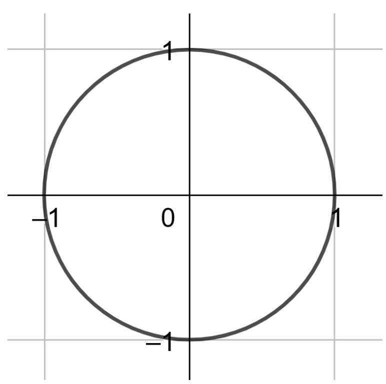


Figure 10.1: Unit Circle for the Euclidean Distance

We also see that lines in the plane are also defined by the distance.

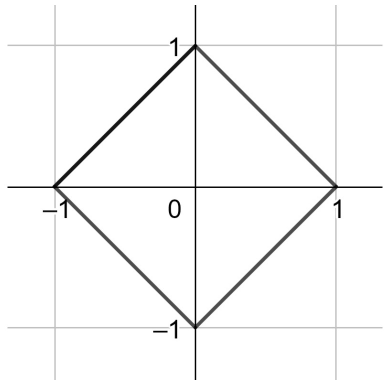


Figure 10.2: Unit Circle for the Euclidean Distance

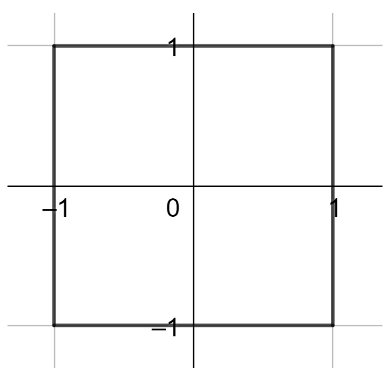


Figure 10.3: Unit Circle for the Euclidean Distance

Definition 10.3. Let d be a distance defined on $\mathbb{R}^2$. A **line** in $\mathbb{R}^2$ is the set of points equidistant from two distinct fixed points,

$$l_{(x,y)} = \left\{ z \in \mathbb{R}^2 \mid d(z,x) = d(z,y) \right\}.$$

For each of the distances on $\mathbb{R}^2$ defined above, the lines are the same.

10.4.2 Related Horizon Content Knowledge

This concept of a metric, or distance, extends beyond the two-dimensional plane. If we look at the real numbers, $\mathbb{R}$, we have the standard metric of $d(x,y) = |x-y|$. We could also define a distance on $\mathbb{R}$ using

$$d_3(x,y) = \frac{|x-y|}{1+|x-y|}.$$

We can also extend the idea of metrics to other sets. If X is a non-empty set the function $d_0 : X \times X \to \{0,1\}$ defined by

$$d_0(x,y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

is called the discrete metric. If we let $C([0,1])$ be the continuous real-valued functions on the closed interval $[0,1]$, we can define a metric on this set as

$$d(f,g) = \max\{f(x) - g(x) \mid x \in [0,1]\}.$$

10.4.3 Exercises

1. Is length (discussed in Section 10.1) different than distance? Explain your answer.
2. Prove that
$$d_3(x,y) = \frac{|x-y|}{1+|x-y|}$$
satisfies the properties for a metric on $\mathbb{R}$.
3. Verify using circles that the lines determined by the two points are the same for Euclidean, Taxicab, and Chebyshev distances.
4. Often distances are given with a direction (e.g., travel three miles due east). For Euclidean distance, find three different ways to give direction.
5. Prove that
$$d(x,y) = \left\{ \begin{array}{ll} 0 & \text{if} \quad x = y \\ 1 & \text{if} \quad x \neq y \end{array} \right.$$
is a metric on the set of real numbers.
6. Prove that
$$d(A,B) = \operatorname{rank}(B-A)$$
is a metric on the set of 2×2 matrices with real coefficients.
7. Using Euclidean distance, find the length of the shortest path from $(-1,7)$ to the x-axis and up to $(8,5)$.

Chapter 11

Groups and Geometry

Erlangen Program

With the rise of non-Euclidean geometries in the 1800's, the mathematics community approached the study of geometry from several different perspectives. Klein [1872] created a new method of characterizing a geometry based on the group structure of the group of transformations, and the properties that are invariant under them. It is named the Erlangen program after the university where Felix Klein was a professor.

11.1 Groups and Transformations

To prepare to study the group of transformations that correspond to Euclidean geometry, we need to do a quick review of group theory.

11.1.1 Group Review

Recall from Section 6.1 the definition of a group:

Definition 11.1 (Group)**.** A non-empty set G, together with a binary operation, $*$, is called a **group** if it satisfies the following conditions:

- $a * b \in G$, $\forall a, b \in G$ (Closure)
- $(a * b) * c = a * (b * c), \forall a, b, c \in G$ (Associative)
- There exists an element $e \in G$ such that for all $a \in G$, $e * a = a * e = a$ (Identity)
- For each $a \in G$, there exists an element $b \in G$ such that $a * b = b * a = e$ (Inverse)

Some of the groups that we have already defined are the real and complex numbers under addition, $(\mathbb{R}, +)$ and $(\mathbb{C}, +)$; the positive real numbers under multiplication, $(\mathbb{R}^+, \cdot)$; the bijections on a set A under function composition; and the 2×2 invertible matrices with determinant of 1 under matrix multiplication, $(SL(2, \mathbb{R}), \cdot)$.

Another group that will prove useful in the study of geometry is the unit circle in the complex plane under multiplication[1],

$$S^1 = \{z \in \mathbb{C} \mid |z| = 1\}.$$

Recall from Section 4.7 that every $z \in S^1$ can be written in the form

$$z = e^{i\theta} = \exp(i\theta) = \cos(\theta) + i\sin(\theta)$$

[1]The notation of S^1 is based on the unit circle being a one-dimensional sphere, where S^2 is the usual two-dimensional sphere in space.

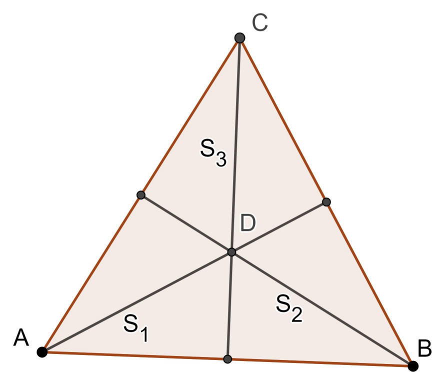
and so we have multiple ways to represent the same complex number, for example $i = \exp\left(i\frac{\pi}{2}\right) = \exp\left(i\frac{5\pi}{2}\right)$. We can see that the group $(S^1, \cdot)$ has a identity of 1 and the inverse of $z = e^{i\theta}$ is $z^{-1} = \overline{z} = e^{-i\theta}$.

11.1.2 Dihedral Groups

Before we look at the transformations of the plane, we will study the transformations of regular polygons.

Let T be an equilateral triangle. Since T is equilateral it has symmetry about the three perpendicular bisectors of the triangle and rotational symmetry about the center of the triangle with angles of 120° and 240°.

We will let r_0 be the identity transformation of the triangle, r_1 be the rotation of the triangle about the center by 120°, and r_2 the be the rotation of the triangle about the center by 240°.



We can also define s_1 to be the reflection over S_1, s_2 be the reflection about S_2, and s_3 be the reflection about S_3. If we consider the operation of composition we can create the following table of these transformations.

$\circ$	r_0	r_1	r_2	s_1	s_2	s_3
r_0	r_0	r_1	r_2	s_1	s_2	s_3
r_1	r_1	r_2	r_0	s_2	s_3	s_1
r_2	r_2	r_0	r_1	s_3	s_1	s_2
s_1	s_1	s_3	s_2	r_0	r_2	r_1
s_2	s_2	s_1	s_3	r_1	r_0	r_2
s_3	s_3	s_2	s_1	r_2	r_1	r_0

We see that this set of transformation under composition forms a group with r_0 being the identity. This group is usually called the dihedral group of symmetries of the triangle, D_3.

We can generalize this to any regular n-gon so that the group of symmetries of an n-gon is labeled D_n. For example, we can represent the 16 elements of D_8 using a stop sign.

11.1.3 Subgroups

As we see in the examples from Chapter 6, groups are often contained in larger groups, $(\mathbb{R}, +) \subset (\mathbb{C}, +)$. We now turn our study to how groups can be nested within each other and the resulting consequences of such nestings by studying subgroups.

Definition 11.2. Let $(G, *)$ be a group and let $S \subseteq G$. Then $(S, *)$ is called a **subgroup** of $(G, *)$ if $(S, *)$ is also a group. We denote this by $S \leq G$, with the binary operation assumed.

From the definition of a subgroup it would appear that one would have to prove all four of the requirements for a set and operation to form a group. However, since we already know information about the set and operation, we already have that the operation is associative on the set and that the larger group has a unique identity and each element of the larger group has a unique inverse. Therefore, proving that a subset can itself be considered a group is much simpler.

Theorem 11.1. *Let $(G, *)$ be a group and $S \subseteq G$. The following are sufficient for determining if $(S, *)$ is a subgroup of $(G, *)$.*

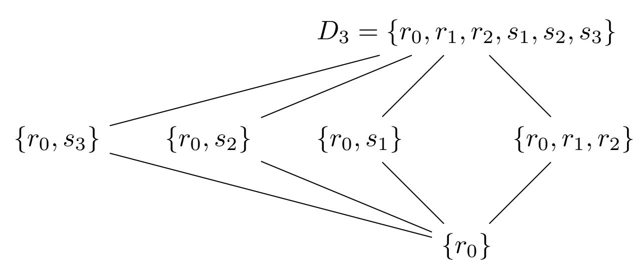
1. *S is closed under the binary operation of G (Closure)*
2. *The identity of G is an element of S (Identity)*
3. *For all $a \in S$, $a^{-1} \in S$. (Inverse)*

We can see that for every group G, the group G itself is a subgroup of G and if e is the identity in G, $\{e\}$ is a subgroup of G, called the **trivial subgroup**. Any subgroup of G that is not the entire group itself is called a **proper subgroup**.

We can use the Cayley table for D_3 to see that the set of rotations, $\{r_0, r_1, r_2\}$, forms a subgroup of D_3 since the composition of two rotations is a rotation, the identity is a rotation, and the inverse of a rotation is a rotation.

We can also see that $(r_0, \circ)$ is a subgroup of D_3. Other subgroups of D_3 include those defined by the subsets $\{r_0, s_1\}, \{r_0, s_2\}, \{r_0, s_3\}$ since $s_i^2 = r_0$ for each of the reflections. As soon as a subgroup has more than one reflection, we can see that the closure condition requires that the subgroup to be the entire group. Similarly, if the subgroup has more than the trivial rotation, it must have all of the rotations.

We see in this discussion that the subgroup structure of D_3 is given by the following diagram.



11.1.4 Exercises

1. Identify whether the given set and operation form a group. If they do not, state which property they fail and give an example.

   a. Integers under addition.
   b. Positive integers under multiplication.
   c. Rational numbers under multiplication.

   d. Even integers under addition.
   e. Odd integers under multiplication.

2. You might notice that none of the operations for groups in this section have been subtraction or division. Why do you think this is? Could a group be formed with the operation of subtraction or division?

3. Let D_4 be the dihedral group of the square.

   a. Create a Cayley table for D_4.
   b. Determine the subgroup structure of D_4.

4. How are the groups $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, and $(\mathbb{R}, +)$ related as subgroups? (Check for subnormal properties.)

5. Let $4\mathbb{Z}$ be the set of integer multiples of 4.

   a. Prove that $(4\mathbb{Z}, +)$ is a subgroup of $\mathbb{Z}$ or provide a counter example.
   b. Prove that $(4\mathbb{Z}, *)$ is a subgroup of $\mathbb{Z}$ or provide a counter example.

6. An alternative way to defined the unit circle is as follows:

$$S^1 = \left\{ z = e^{i\theta} \mid \theta \in \mathbb{R} \right\}.$$

   Given that we know $\mathbb{Z}$ forms a group under addition, show that $K = \left\{ z = e^{i\theta} \mid k \in \mathbb{Z} \right\}$ is a subgroup of S^1. You do not need to show associativity.

11.2 Normal Subgroups and Factor Groups

11.2.1 Cosets

Definition 11.3. Let S be a subgroup of G. The subset $aS = \{ab : b \in S\}$ of G is the **left coset of** S **containing** a and $Sa = \{ba : b \in S\}$ is the **right coset of** S **containing** a.

Theorem 11.2. *For a group G with identity e, for each element $a \in G$ + $aG = Ga = G$ + $a\{e\} = \{e\}a = \{a\}$*

Proof. In order to prove that the left and right cosets of G containing a are always the entire group G, we choose a generic element of G and we can call it b. Since G is a group, $a^{-1}b \in G$ and so $b = a(a^{-1}b) \in aG$. Similarly, since $ba^{-1} \in G$, $b = (ba^{-1})a \in Ga$. So $aG = Ga = G$.

Since $ae = a$, we can also see that $a\{e\} = \{a\}$ and $\{e\}a = \{a\}$. □

If we return to the group D_3, we can look at the left and right cosets related to the corresponding non-trivial subgroups. If we let $S = \{r_0, s_1\}$, we have the following left and right cosets,

Left Cosets of $\{r_0, s_1\}$	Right Cosets of $\{r_0, s_1\}$
$r_0 \circ S = \{r_0, s_1\} = S$	$S \circ r_0 = \{r_0, s_1\} = S$
$r_1 \circ S = \{r_1, s_3\}$	$S \circ r_1 = \{r_1, s_2\}$
$r_2 \circ S = \{r_2, s_2\}$	$S \circ r_2 = \{r_2, s_3\}$
$s_1 \circ S = \{s_1, r_0\} = S$	$S \circ s_1 = \{s_1, r_0\} = S$
$s_2 \circ S = \{s_2, r_2\}$	$S \circ s_2 = \{s_2, r_1\}$
$s_3 \circ S = \{s_3, r_1\}$	$S \circ s_3 = \{s_3, r_2\}$

and we see that many of the left cosets are different from the corresponding right cosets. On the other hand, if we let $S = \{r_0, r_1, r_2\}$ we have the following left and right cosets.

Left Cosets of $\{r_0, r_1, r_2\}$	Right Cosets of $\{r_0, r_1, r_2\}$
$r_0 \circ S = \{r_0, r_1, r_2\} = S$	$S \circ r_0 = \{r_0, r_1, r_2\} = S$
$r_1 \circ S = \{r_1, r_2, r_3\} = S$	$S \circ r_1 = \{r_1, r_2, r_3\} = S$
$r_2 \circ S = \{r_2, r_0, r_1\} = S$	$S \circ r_2 = \{r_2, r_0, r_1\} = S$
$s_1 \circ S = \{s_1, s_2, s_3\}$	$S \circ s_1 = \{s_1, s_3, s_2\}$
$s_2 \circ S = \{s_2, s_3, s_1\}$	$S \circ s_2 = \{s_2, s_1, s_3\}$
$s_3 \circ S = \{s_3, s_1, s_2\}$	$S \circ s_3 = \{s_3, s_2, s_1\}$

In this situation, we see that the left cosets and the right cosets correspond and that the group D_3 is partitioned into two sets, $\{r_0, r_1, r_2\}$ and $\{s_1, s_2, s_3\}$ by the two cosets.

11.2.2 Normal Subgroups

The subgroups for which the left and right cosets correspond are called normal.

Definition 11.4. A subgroup N of a group G is called a **normal subgroup** if it is invariant under conjugation, that is if $gN = Ng$ for all $g \in G$. We denote this by $N \lhd G$.

Theorem 11.3. *A subgroup N of a group G is a normal subgroup if and only if for all $g \in G$ and $f \in N$, $g^{-1} * f * g \in N$.*

Proof. Let's first assume that N is a normal subgroup of G. Then for all $g \in G$, $gN = Ng$. If $f \in N$, we see that $fg \in gN$ and so there exists $\tilde{f} \in N$ such that $fg = g\tilde{f}$ and so $g^{-1}fg \in N$. Since f and g were arbitrary, $g^{-1} * f * g \in N$ for all $g \in G$ and $f \in N$.

Let's alternatively assume that $g^{-1} * f * g \in N$ for all $g \in G$ and $f \in N$. Then if we choose an element $g \in G$, we know that $gN = \{gf \mid f \in N\}$ and $Ng = \{fg \mid f \in N\}$. If $fg \in Ng$, we know that $g^{-1}fg \in N$ and so there exists $\tilde{f} \in N$ such that $g^{-1}fg = \tilde{f}$ and $fg = g\tilde{f}$, implying that $fg \in gN$. So $Ng \subseteq gN$. We can similarly show that $gN \subseteq Ng$ and so $gN = Ng$ and N is subnormal. ☐

Theorem 11.4. *If N is a normal subgroup of G, then for every $n \in N$, $nN = N$.*

11.2.3 Factor Groups

If N is a normal subgroup of a group G, we can create a relation on G as follows: For $g, h \in G$

$$g \sim h \Leftrightarrow gN = hN.$$

Since $gN = gN$, we see that the relation is reflexive. Since $gN = hN$ is equivalent to $hN = gN$ based on properties of sets we see that the relation is symmetric. In order to show that the relation is transitive, we assume that $g, h, m \in G$ such that $g \sim h$ and $h \sim m$ and see that transitivity of set equality implies that $g \sim m$ and so the relation is transitive.

Since this relation is an equivalence relation we see that it creates a partition of G into cosets. We can then define

$$\frac{G}{N} = \{gN \mid g \in G\}.$$

We can also define an operation on this set using the operation from G as

$$(gN)(hN) := (gh)N.$$

To verify that this operation is well defined, we can choose two other representations of these cosets so that $gN = g_1N$ and $hN = h_1N$ and verify that the product of the cosets is the same independent of the representation. Since $gN = g_1N$, there exists $n_1 \in N$ such that $gn_1 = g_1$, and since $hN = h_1N$, there exists $n_2 \in N$ such that $hn_2 = h_1$. Therefore,

$$(g_1h_1)N = (gn_1)(hn_2)N = g(n_1h)N$$

since $n_2N = N$. And since N is normal, there exist $n_3 \in N$ such that $n_1h = hn_3$, so

$$(g_1h_1)N = (gh)n_1N = (gh)N.$$

So the operation is well-defined.

Theorem 11.5. *If N is a normal subgroup of G, then $\frac{G}{N}$ with the operation inherited from G is a group.*

This group $\frac{G}{N}$ is called the quotient group of G with respect to N.

11.2.4 Exercises

1. Consider the relationship between $(\mathbb{Z},+)$ and $(\mathbb{R},+)$.

   a. Show that $(\mathbb{Z},+)$ is a normal subgroup of $(\mathbb{R},+)$.
   b. Describe the elements of the factor group

   $$\frac{(\mathbb{R},+)}{(\mathbb{Z},+)}$$

   and the operation induced from $(\mathbb{R},+)$.

2. If a group is abelian, what does this tell us about normality of the subgroups? Why?

3. Prove that the intersection of normal subgroups of a group G is again a normal subgroup of G.

11.3 Group Homomorphisms Revisited

Recall the following definition of group homomorphism and isomorphism.

Definition 11.5. A map, ϕ, from a group $(G,*)$ to a group $(G',\cdot)$ is called a **homomorphism** if

$$\phi(a*b) = \phi(a)\cdot\phi(b)$$

for every $a,b \in G$. If ϕ is also a bijection, then we say that ϕ is an **isomorphism**.

Now that we know when two groups are isomorphic to one another, we turn to looking at how “different” two groups are that are not isomorphic, but have a homomorphism between them by looking at the set of points of the domain that are mapped to the identity. This provides information about how far away the group homomorphism is from a one-to-one function .

Theorem 11.6. *Let $\phi : G \to G'$ be a group homomorphism. The set $\{a \in G : \phi(a) = e'\}$, where e' is the identity in G', is a subgroup of G and is called the kernel of ϕ, denoted by Ker(ϕ).*

Proof. Since ϕ is a group homomorphism, if e is the identity in G, $\phi(a) = \phi(a*e) = \phi(a)*\phi(e)$ for all $a \in G$. So $\phi(e)$ is the identity in G'. Therefore, $e \in \mathrm{Ker}(\phi)$.

If $a,b \in \mathrm{Ker}(\phi)$, then

$$\phi(a*b) = \phi(a)\cdot\phi(b) = e'\cdot e' = e'$$

and so $a*b \in \mathrm{Ker}(\phi)$.

If $a \in \text{Ker}(\phi)$, then

$$\phi(a^{-1}) = \phi(a^{-1}) \cdot e' = \phi(a^{-1}) \cdot \phi(a) = \phi(a^{-1} * a) = \phi(e) = e'$$

and so $a^{-1} \in \text{Ker}(\phi)$.

Therefore, $\text{Ker}(\phi)$ is a subgroup of G. $\square$

Now that we know that the kernel of a group homomorphism is a subgroup, we prove that it is in fact a normal subgroup.

Theorem 11.7. *If $\phi : G \to G'$ is a group homomorphism, then $Ker(\phi) \lhd G$.*

Proof. By the previous theorem, we have proven that $\text{Ker}(\phi)$ is a subgroup of G. Thus it is sufficient to prove that

$$g\text{Ker}(\phi) = \text{Ker}(\phi)g \quad \forall g \in G.$$

Let $g \in G$ and let $a \in \{x \in G : \phi(x) = \phi(g)\}$. Then since $\phi(a) = \phi(g)$, we can multiply both sides of the equation by $\phi(g)^{-1}$ on the left (respectively right) and we have that

$$\phi(g)^{-1}\phi(a) = e' \quad \left(\text{respectively } \phi(a)\phi(g)^{-1} = e'\right).$$

Since ϕ is a homomorphism, this is equivalent to

$$\phi(g^{-1}a) = e' \quad \left(\text{repectively } \phi(ag^{-1}) = e'\right).$$

So $g^{-1}a$ and ag^{-1} are elements of $\text{Ker}(\phi)$. Therefore, since $a = g(g^{-1}a)$ and $a = (ag^{-1})g$, we have that $a \in g\text{Ker}(\phi)$ and $a \in \text{Ker}(\phi)g$ and so

$$\{x \in G : \phi(x) = \phi(g)\} \subseteq g\text{Ker}(\phi) \quad \text{and} \quad \{x \in G : \phi(x) = \phi(g)\} \subseteq \text{Ker}(\phi)g$$

Let $a \in g\text{Ker}(\phi)$. Then there exists a $b \in \text{Ker}(\phi)$ such that $a = gb$. Thus $b = g^{-1}a$ and so $e' = \phi(b) = \phi(g^{-1}a) = \phi(g)^{-1}\phi(a)$ and so $\phi(a) = \phi(g)$. Similarly if $a \in \text{Ker}(\phi)$, then $\phi(a) = \phi(g)$.

Therefore,

$$g\text{Ker}(\phi) = \{x \in G : \phi(x) = \phi(g)\} = \text{Ker}(\phi)g$$

and so $\text{Ker}(\phi) \lhd G$. $\square$

So we have shown that for every homomorphism, the kernel of the homomorphism is a normal subgroup. Alternatively, if we start with a normal subgroup, we can create a homomorphism for which the normal subgroup is equal to the kernel.

Theorem 11.8. *Let G be a group and let N be a normal subgroup of G. Then the function $\gamma : G \to \frac{G}{N}$ defined by $\gamma(g) = gN$ is a group homomorphism and $Ker(\gamma) = N$.*

Proof. Because

$$\gamma(g_1g_2) = (g_1g_2)N = (g_1N)(g_2N) = \gamma(g_1)\gamma(g_2)$$

we know that γ is a group homomorphism.

Since the only elements $g \in G$ such that $gN = N$ are the elements of N, we have that $\text{Ker}(\gamma) = N$. $\square$

This means that there is a strong relationship between normal subgroups and kernels of group homomorphisms. This leads us to one of the most important theorems in basic group theory.

Theorem 11.9 (Fundamental Homomorphism Theorem)**.** *Let $\phi : G \to G'$ be a group homomorphism with kernel Ker(ϕ). Then $\phi(G)$ is a group and the map*

$$\hat{\phi} : \frac{G}{Ker(\phi)} \to \phi(G)$$

defined by $\hat{\phi}\,(aKer(\phi)) = \phi(a)$ is an isomorphism.

Proof. We first need to prove that $\phi(G)$ is a subgroup of G'. Since ϕ is a homomorphism, $\phi(G)$ is closed under the binary operation. From Theorems 6.3 and 6.4 we have that the identity of G' is in $\phi(G)$ and if $\phi(a) \in \phi(G)$ then $\phi(a)^{-1} \in \phi(G)$. So $\phi(G)$ is a subgroup of G'.

If a and a' are two elements of G such that $a\mathrm{Ker}(\phi) = a'\mathrm{Ker}(\phi)$, we know that there are elements b and b' in $\mathrm{Ker}(\phi)$ such that $ab = a'b'$. So $a' = ab(b')^{-1}$ and

$$\phi(a') = \phi(ab(b')^{-1}) = \phi(a)\phi(b)\phi(b')^{-1} = \phi(a)$$

so $\hat{\phi}$ is well-defined.

If $a\mathrm{Ker}(\phi)$ and $b\mathrm{Ker}(\phi)$ then

$$\hat{\phi}\,((a\mathrm{Ker}(\phi))\,((b\mathrm{Ker}(\phi))) = \hat{\phi}\,((ab)\mathrm{Ker}(\phi)) = \phi(ab) = \phi(a)\phi(b) = \hat{\phi}\,(a\mathrm{Ker}(\phi))\,\hat{\phi}\,(b\mathrm{Ker}(\phi))$$

and so $\hat{\phi}$ is a homomorphism.

If $a\mathrm{Ker}(\phi)$ is in the kernel of $\hat{\phi}$, we know that $\phi(a)$ is the identity of G' and so $a \in \mathrm{Ker}(\phi)$. This implies that $a\mathrm{Ker}(\phi) = \mathrm{Ker}(\phi)$ and so $\mathrm{Ker}(\hat{\phi}) = \mathrm{Ker}(\phi)$ and so $\hat{\phi}$ is one-to-one. By the definition of $\phi(G)$ we know that $\hat{\phi}$ is onto and so is a bijection.

Therefore, $\hat{\phi}$ is an isomorphism. $\square$

The prior two theorems can be summarized in the following commutative diagram.

$$\begin{tikzcd} G \arrow[r, "\phi"] \arrow[d, "\gamma"'] & G' \\ \frac{G}{\mathrm{Ker}(\phi)} \arrow[ur, "\hat{\phi}"'] & \end{tikzcd}$$

11.3.1 Exercises

1. Use the Fundamental Homomorphism Theorem to prove that

$$\frac{\mathbb{R}}{\{2\pi n | n \in \mathbb{Z}\}} \cong \{z \in \mathbb{C} \mid |z| = 1\}$$

2. Use the Fundamental Homomorphism Theorem to prove that

$$(\mathbb{R}, +) \cong (\mathbb{R}^+, \cdot)$$

Chapter 12

Euclidean Transformational Geometry

12.1 Introduction to Transformational Geometry

The axiomatic systems of Chapter 9 define objects as being the 'same', or congruent, if the measures of the lengths and angles are the same. The remainder of the geometric system is derived from this concept of congruence and the other axioms. Another approach is to axiomatically define a set and a distance on the set. We then define a group of functions on this set that maintain distance, called isometries, and define objects to be congruent if there is an isometry that maps one object onto the other object.

12.1.1 Neutral Geometry from Distances

Before we return to the study of Euclidean geometry from the perspective of transformations, we will study properties of geometry that are independent of a specific set and distance.

Throughout this section we will let X be a non-empty set with a distance d. We now need to define some of the basic geometric objects in terms of the distance d and the related concept of betweenness.

Definition 12.1. A **line segment** is the union of two distinct points, A and B, and all the points **between** those two points, denoted as

$$\overline{AB} = \{C \in X \mid d(A,C) + d(C,B) = d(A,B)\}\,.$$

Definition 12.2. A **ray**, $\overrightarrow{AB}$, is the union of the segment $\overline{AB}$ and the set of all points C such that B is between A and C,

$$\overrightarrow{AB} = \{C \in X \mid d(A,C) + d(C,B) = d(A,B)\} \cup \{C \in X \mid d(A,B) + d(B,C) = d(A,C)\}\,.$$

Note that the ray $\overrightarrow{AB}$ is distinct from the ray $\overrightarrow{BA}$, with their intersection being the segment $\overline{AB}$. When we take the union of the two rays we get a line.

Definition 12.3. For two points, A and B, the **line** containing A and B, $\overleftrightarrow{AB}$, is the union of the rays $\overrightarrow{AB}$ and $\overrightarrow{BA}$.

When the set is the plane with the Euclidean distance, the definition given here is equivalent to the set of points equidistant from two fixed points. If we consider the set to be $\mathbb{R}^3$ with the standard Euclidean distance, this definition of a line is distinct from the other since the points equidistant from two fixed points in that setting is a plane.

Definition 12.4. An **angle** is the union of two noncollinear rays with a common endpoint. The common endpoint is called the vertex of the angle, and the rays are called the sides of the angle.

As we saw in Chapter 9, properties of triangles are critical in the development and so we need to have a definition of a triangle for these spaces.

Definition 12.5. A **triangle** is the union of three segments determined by three noncollinear points; for three noncolinear points A, B, and C,

$$\triangle ABC = \overline{AB} \cup \overline{BC} \cup \overline{CA}.$$

Each of the three noncollinear points that determine a triangle is called a **vertex** of the triangle.

Definition 12.6. A **circle** is the set of points equidistant from a fixed point.

12.1.2 Isometries

We will continue to let X be a set with distance d so that the results can be applied in a more general setting.

Definition 12.7. A function $f : X \to X$ is called an **isometry**, or **rigid motion**, if it is a surjection that maintains distance between any two points,

$$d(f(x), f(y)) = d(x, y), \quad \forall x, y \in X.$$

While we have defined isometries to be surjections[1], we also see that they are injections as a result of the distance maintaining property.

Theorem 12.1. *Isometries are one-to-one.*

Proof. Let (X, d) be a set and a distance on the set and let $f : X \to X$ be an isometry. If we assume that $f(x) = f(y)$, we know from the identity of indiscernibles for a distance that $d(f(x), f(y)) = 0$. Since f is an isometry,

$$d(x, y) = d(f(x), f(y)) = 0$$

and so we know that $x = y$. Therefore, f is an injection. □

Therefore, isometries are bijections. From Section 5.2 we know that compositions of bijections are bijections and bijections are invertible with the inverse function also being a bijection.

Theorem 12.2. *If f and g are isometries, then $g \circ f$ is also an isometry.*

Proof. Since we already know that compositions of bijections are bijections, it suffices to prove the properties of maintaining distances. Let $f : X \to X$ and $g : X \to X$ be isometries. Then for any $x, y \in X$,

$$d\left((g \circ f)(x), (g \circ f)(y)\right) = d\left(g(f(x)), g(f(y))\right) = d\left(f(x), f(y)\right)$$

because g is an isometry. Since f is also an isometry, $d\left(f(x), f(y)\right) = d(x, y)$ and

$$d\left((g \circ f)(x), (g \circ f)(y)\right) = d(x, y).$$

Therefore, $g \circ f$ is an isometry. □

Theorem 12.3. *If f is an isometry, then f is invertible and f^{-1} is also an isometry.*

[1] The assumption of surjectivity is not needed for Euclidean spaces due to completeness properties of the real numbers. This assumption is included here to avoid those details.

Proof. Since f is an isometry, it is also a bijection. This means that f^{-1} exists. If we let x and y be elements of X, since f is a surjection, there exist $a, b \in X$ such that $x = f(a)$ and $y = f(b)$. So

$$d(f^{-1}(x), f^{-1}(y)) = d(f^{-1}(f(a)), f^{-1}(f(b)) = d(a, b) = d(f(a), f(b)) = d(x, y)$$

since f is an isometry. □

Theorem 12.4. *Betweenness of points is invariant under an isometry.*

Proof. Let f be an isometry. Let A, B, and C be three distinct points such that B is between A and C. We will let $A' = f(A)$, $B' = f(B)$, and $C' = f(C)$. By the definition of betweenness of points, $d(A, C) = d(A, B) + d(B, C)$ and A, B, and C are collinear. Since f is an isometry,

$$d(A', C') = d(A, C) = d(A, B) + d(B, C) = d(A', B') + d(B', C').$$

Thus, by the Triangle Inequality, A', B', and C' are collinear. Therefore, B' is between A' and C'. □

Since line segments, rays, and angles are all defined by properties of betweenness, we see that isometries maintain these properties. Since triangles are defined by segments, triangles are also maintained by isometries.

Corollary 12.1. *The image of a line segment (ray, angle, or triangle) under an isometry of the plane is a line segment (ray, angle, or triangle).*

Since isometries maintain distance, the length of segments are also maintained under the transformation.

Corollary 12.2. *The image of a line segment under an isometry of the plane is a line segment with the same length, distance between endpoints, as the original.*

Because the length of the sides of a triangle are all maintained under an isometry, we see that isometries maintain triangle properties.

Corollary 12.3. *The image of a triangle under an isometry of the plane is a triangle whose sides have the same lengths as the original.*

Theorem 12.5. *Isometries map lines to lines and circles to circles.*

Related Content Standards

- (8.G.1) Verify experimentally the properties of rotations, reflections, and translations:
  a. Lines are taken to lines, and line segments to line segments of the same length.
  b. Angles are taken to angles of the same measure.
  c. Parallel lines are taken to parallel lines.

12.1.3 Congruence

We now define the property of congruence based on isometries.

Definition 12.8. Two planar objects are said to be **congruent** if there is an isometry that maps one onto the other.

Based on the properties of bijections and isometries, we can show that congruence is an equivalence relation.

Related Content Standards

- (HSG.CO.7) Use the definition of congruence in terms of rigid motions to show that two triangles are congruent if and only if corresponding pairs of sides and corresponding pairs of angles are congruent.
- (HSG.CO.8) Explain how the criteria for triangle congruence (ASA, SAS, and SSS) follow from the definition of congruence in terms of rigid motions.

Theorem 12.6. *Two triangles are congruent if and only if corresponding pairs of sides and corresponding pairs of angles are congruent.*

12.1.4 Exercises

1. Prove that congruence satisfies the properties of an equivalence relation.
2. Prove that two circles are congruent if and only if they have the same radii, and hence the same area.

12.2 Representations of the Euclidean Plane

As we study transformational geometry of the plane we will use four distinct perspectives.

12.2.1 Synthetic Plane

The synthetic perspective follows the methods of Euclid without the use of coordinates or formulas. With this synthetic approach, the concept of distance and measurement is always with respect to a separately defined unit.

12.2.2 Cartesian Plane

In the 1600s René Descartes and Pierre de Fermat independently developed the use of a coordinate plane to provide an analytic foundation for the study of geometry. While some may consider the synthetic methods to be more axiomatic, we have shown in Chapter 4 that the construction of the real number system, and thus also $\mathbb{R}^2$, is just as axiomatic using the axioms of set theory. When we follow the real analytic perspective of transformational geometry we will generally use the distance defined by the Euclidean distance,

$$d\left((x,y),(w,z)\right) = \sqrt{(w-x)^2+(z-y)^2},$$

and we will consider the transformations as $f : \mathbb{R}^2 \to \mathbb{R}^2$ with

$$f(x,y) = (f_1(x,y), f_2(x,y))$$

where f_1 and f_2 are functions from $\mathbb{R}^2$ to $\mathbb{R}$.

Furthermore, in $\mathbb{R}^2$ we see that lines can be expressed in the form

$$\left\{(x,y) \in \mathbb{R}^2 \mid y = mx + b \text{ for some } m, b \in \mathbb{R}\right\}$$

and circles centered at the point (x_0, y_0) with a radius of r can be expressed as

$$\left\{(x,y) \in \mathbb{R}^2 | (x-x_0)^2 + (y-y_0)^2 = r^2\right\}.$$

12.2.3 Vector Space

An alternative perspective of the plane involves viewing $\mathbb{R}^2$ as a two-dimensional real vector space. From this perspective, we define

$$\mathbb{R}^2 = \{\langle x, y\rangle \mid x, y \in \mathbb{R}\}$$

with the inner product defined as

$$\langle x, y\rangle \cdot \langle w, z\rangle = xw + yz,$$

and the norm of a vector defined as

$$||\langle x, y\rangle|| = \sqrt{\langle x, y\rangle \cdot \langle x, y\rangle} = \sqrt{x^2 + y^2}.$$

We define vector addition defined by $\langle x, y\rangle + \langle w, z\rangle = \langle x+w, y+z\rangle$, scalar multiplication by $c\langle x, y\rangle = \langle cx, cy\rangle$, and so $\langle x, y\rangle - \langle w, z\rangle = \langle x - w, y - z\rangle$ and we can define the distance between vectors as

$$d\left(\langle x, y\rangle, \langle w, z\rangle\right) = ||\langle x, y\rangle - \langle w, z\rangle||.$$

We will focus much of our attention on linear transformations of this vector space represented by matrices,

$$T_A\left(\langle x, y\rangle\right) = A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} = \langle ax + by, cx + dy\rangle.$$

12.2.4 Complex Plane

In the 1700s Leonard Euler proposed the concept of considering the complex numbers as isomorphic to the real plane,

$$\mathbb{C} = \left\{x + iy \mid (x, y) \in \mathbb{R}^2\right\}$$

with distance

$$d(w, z) = |w - z|.$$

Then transformations of the plane are simply functions of the form $f : \mathbb{C} \to \mathbb{C}$. The circle centered at the complex number c with radius r in $\mathbb{C}$ can be expressed as

$$C_{(c,r)} = \{z \in \mathbb{C} \mid |z - c| = r\}.$$

The details of the complex plane perspective are in Section 4.7.

12.2.5 Isometries

Recall that the definition of an isometry is a surjection that maintains distances between any two points. We will see that the uses of the different representations of the plane will help us to better understand and categorize these isometries.

Related Content Standards

- (HSG.CO.2) Represent transformations in the plane using, e.g., transparencies and geometry software; describe transformations as functions that take points in the plane as inputs and give other points as outputs. Compare transformations that preserve distance and angle to those that do not (e.g., translation versus horizontal stretch).

The representation of the plane using complex numbers contains a significant amount of information. Using the properties of the complex numbers we can determine which functions $f : \mathbb{C} \to \mathbb{C}$ are isometries.

Theorem 12.7. *$f : \mathbb{C} \to \mathbb{C}$ is an isometry if and only if*

$$f(z) = az + b \quad \text{or} \quad f(z) = a\overline{z} + b$$

for some $a, b \in \mathbb{C}$ with $|a| = 1$.

Proof. If f is an isometry, then

$$|f(z) - f(w)| = |z - w|$$

for all $z, w \in \mathbb{C}$. In order to better understand these isometries we will first consider some simpler isometries.

Assume that $g : \mathbb{C} \to \mathbb{C}$ is an isometry such that $g(0) = 0$ and $g(1) = 1$. Then the properties of isometries implies that for all complex numbers z,

$$|g(z)| = |z| \quad \text{and} \quad |g(z) - 1| = |z - 1|.$$

Squaring each of these equations gives us the equivalent equations

$$g(z)\overline{g(z)} = z\overline{z} \quad \text{and} \quad (g(z) - 1)\overline{(g(z) - 1)} = (z - 1)\overline{(z - 1)}.$$

Using properties of complex conjugation and distribution the second equation can be rewritten as

$$g(z)\overline{g(z)} - (g(z) + \overline{g(z)}) + 1 = z\overline{z} - (z + \overline{z}) + 1$$

and substituting the first equation we have that

$$g(z) + \overline{g(z)} = z + \overline{z}.$$

This means that $\mathrm{Re}(g(z)) = \mathrm{Re}(z)$. Since $|g(z)| = |z|$ we can infer that $\mathrm{Im}(g(z)) = \pm\mathrm{Im}(z)$ and so $g(z) = z$ or $g(z) = \overline{z}$.

If $f : \mathbb{C} \to \mathbb{C}$ is a generic isometry, we can let $b = f(0)$ and see that $h : \mathbb{C} \to \mathbb{C}$ defined by $h(z) = f(z) - b$ is an isometry such that $h(0) = 0$. This would then imply that $|h(1)| = |h(1) - h(0)| = |1|$. We can then let $a = h(1)$ and we see that $|a| = 1$ and so $a^{-1} = \overline{a}$. If we let $g : \mathbb{C} \to \mathbb{C}$ be defined by $g(z) = \overline{a}h(z) = \overline{a}(f(z) - b)$ we see that g is an isometry with $g(0) = 0$ and $g(1) = 1$. So by the argument in the previous paragraph, $g(z) = z$ or $g(z) = \overline{z}$. Therefore,

$$f(z) = az + b \quad \text{or} \quad f(z) = a\overline{z} + b$$

for some $a, b \in \mathbb{C}$ with $|a| = 1$. □

Over the next several sections we will see how the properties of these expressions correspond to the transformations of translations, rotations, reflections, and glide-reflections.

Definition 12.9. Let I be the set of isometries of the plane.

$$I = \{f : \mathbb{C} \to \mathbb{C} \mid f(z) = az + b \text{ or } f(z) = a\overline{z} + b \text{ for some } a, b \in \mathbb{C} \text{ with } |a| = 1\}$$

12.2.6 Exercises

1. For each of the following functions, $f : \mathbb{R}^2 \to \mathbb{R}^2$, determine if f is an isometry. Justify your answer.

   a. $f(x, y) = (x + 2, y - 3)$
   b. $f(x, y) = (x^2, y)$
   c. $f(x, y) = (-x, -y)$
   d. $f(x, y) = \left(\dfrac{x + \sqrt{3}y}{2}, \dfrac{-\sqrt{3}x + y}{2}\right)$
   e. $f(x, y) = (x, -y)$

2. If $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ is an isometry represented by a 2×2 matrix, A, as

$$T_A(\langle x, y\rangle) = A\begin{pmatrix} x \\ y \end{pmatrix},$$

what can we say about the determinant of A?

3. For each of the following functions, $f : \mathbb{C} \to \mathbb{C}$, determine if f is an isometry. Justify your answer by comparing $|f(z) - f(w)|$ and $|z - w|$ for generic complex numbers z and w.

   a. $f(z) = z + 2 - 3i$
   b. $f(z) = z^2$
   c. $f(z) = \overline{z} + 2i$
   d. $f(z) = 3\overline{z}$
   e. $f(z) = e^{i\frac{\pi}{4}} z + i$

4. Use the form of isometries for the complex plane,

$$I = \{f : \mathbb{C} \to \mathbb{C} \mid f(z) = az + b \text{ or } f(z) = a\overline{z} + b \text{ for some } a, b \in \mathbb{C} \text{ with } |a| = 1\},$$

to prove each of the theorems in the previous section about isometries.

   a. If f and g are isometries, then $g \circ f$ is an isometry.
   b. If f is an isometry, then f^{-1} is invertible and f^{-1} is also an isometry.
   c. Isometries map lines to lines and circles to circles.

5. Use the form of isometries for the complex plane,

$$I = \{f : \mathbb{C} \to \mathbb{C} \mid f(z) = az + b \text{ or } f(z) = a\overline{z} + b \text{ for some } a, b \in \mathbb{C} \text{ with } |a| = 1\},$$

to show that I, together with the operation of function composition, is a group.

12.3 Translations

Our first type of isometry that we discuss are translations that involve a shifting of the plane, without any turning. If we use a sheet of paper to represent the plane, these transformations are a sliding of the paper without changing the orientation of the paper. Intuitively we see that this really does not change the plane at all, it just changes our perspective of the plane.

12.3.1 Synthetic Plane

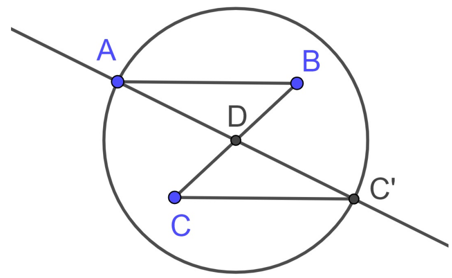
Definition 12.10. Given a segment $\overline{AB}$, we define the translation $T_{\overline{AB}}$ for every point C by $C' = T_{\overline{AB}}(C)$ is the unique point in the plane so that $\overline{CC'}$ has the same length as $\overline{AB}$ and is parallel to $\overline{AB}$.

So if we start with a segment $\overline{AB}$ and a point C, we can find the midpoint of the segment $\overline{BC}$ and label it as D. Then we find the point on the line $\overleftrightarrow{AD}$ that is the same distance from D as A, and label the point C'.

If C was not on the line $\overleftrightarrow{AB}$, $\overline{BC}$ and $\overline{AC'}$ bisect each other, and so $ABC'C$ is a parallelogram and so $\overline{CC'}$ is the same length as, and parallel to, $\overline{AB}$.

If C was on the line $\overline{AB}$, we see that C' is also on the same line and the length of $\overline{CC'}$ is the same as $\overline{AB}$. So we see that the translations on the synthetic plane are constructible.

To prove that these translations are isometries we let C and D be any two points of the plane and we let $C' = T_{\overline{AB}}(C)$ and $D' = T_{\overline{AB}}(D)$. Since $\overline{CC'}$ and $\overline{DD'}$ are two segments of the same length and are parallel to one another, properties of parallelograms determine that $CDD'C'$ is a parallelogram and $\overline{CD}$ is the same length as $\overline{C'D'}$. So the translation maintains the distance between points on the plane.



Related Content Standards

- (HSG.CO.4) Develop definitions of rotations, reflections, and translations in terms of angles, circles, perpendicular lines, parallel lines, and line segments.

12.3.2 Cartesian Plane and Vector Space

When we add the coordinates to the plane, the definition and properties of the translation become easier to describe due to the increased amount of information provided by the Cartesian coordinate system. In particular, since any translation can be viewed as a combination of horizontal and vertical translations we see that a translation of a set length in a certain direction is defined by these two simpler translations.

Definition 12.11. Given a point $(h,k) \in \mathbb{R}^2$, we define $T_{(h,k)}$ by

$$T_{(h,k)}\left((x,y)\right) = (x+h, y+k).$$

A translation corresponding to a line segment between the points $A = (a_1, a_2)$ and $B = (b_1, b_2)$ in this perspective corresponds to the point (h,k), where $h = b_1 - a_1$ and $k = b_2 - a_2$. We can also show that translations are isometries using this perspective by letting $C = (c_1, c_2)$ and $D = (d_1, d_2)$ be two points in $\mathbb{R}^2$ and let (h,k) be a point in $\mathbb{R}^2$. Then

$$\begin{aligned} d\left(T_{(h,k)}(C), T_{(h,k)}(D)\right) &= d\left((c_1+h, c_2+k), (d_1+h, d_2+k)\right) \\ &= \sqrt{\left((d_1+h)-(c_1+h)\right)^2 + \left((d_2+k)-(c_2+k)\right)^2} \\ &= \sqrt{(d_1-c_1)^2 + (d_2-c_2)^2} = d\left(C, D\right) \end{aligned}$$

and so $T_{(h,k)}$ maintains distances.

Using a basic understanding of vectors, the properties of translations in the Cartesian perspective translate very easily to the vector space perspective.

Definition 12.12. Given a vector $\langle h,k\rangle$, we define $T_{\langle h,k\rangle}$ by

$$T_{\langle h,k\rangle}\left(\langle x,y\rangle\right) = \langle x,y\rangle + \langle h,k\rangle = \langle x+h, y+k\rangle.$$

12.3.3 Complex Plane

Complex numbers provide even more structure than $\mathbb{R}^2$ and so the notation for translations is simplified.

Definition 12.13. Given a complex number $b \in \mathbb{C}$, we define T_b by

$$T_b(z) = z + b.$$

This additional structure provided by $\mathbb{C}$ also simplifies the proof that translations are isometries,

$$|T_b(w) - T_b(z)| = |(w+b) - (z+b)| = |z - w|.$$

12.3.4 Compositions

Theorem 12.8. *The composition of two translations is a translation.*

Proof. Let T_b and T_c be two translations of $\mathbb{C}$. Then

$$T_b \circ T_c(z) = T_b(z+c) = (z+c)+b = z+(c+b) = T_{c+b}(z).$$

□

We can also see that the order of the translations does not matter.

12.3.5 Group of Translations

Theorem 12.9. *Let T be the set of translations of the plane. Then with the operation of function composition, $(T, \circ)$ is a subgroup of $(I, \circ)$.*

Proof. Since the composition of two translations is also a translation, $(T, \circ)$ is closed.

Since the identity function can be considered a translation, $T_0(z) = z + 0$, we see that $(T, \circ)$ has an identity.

Since for any $b \in \mathbb{C}$, $T_b \circ T_{-b}(z) = z - b + b = z$, we see that T_b is invertible and $T_b^{-1} = T_{-b}$. So the inverse of any translation is also a translation.

Therefore, $(T, \circ)$ is a subgroup of $(I, \circ)$. □

Recall from Section 6.1 that two groups, $(G, *)$ and $(G', \times)$, are isomorphic ('equivalent') if there is a bijection $\phi : G \to G'$ such that

$$\phi(a * b) = \phi(a) \times \phi(b)$$

for all $a, b \in G$.

If we let $\phi : \mathbb{C} \to T$ be defined by $\phi(b) = T_b$, we see that

$$\phi(b+c) = T_{b+c} = T_b \circ T_c = \phi(b) \circ \phi(c),$$

making ϕ a group homomorphism. From the definition of translations on $\mathbb{C}$ we can see that ϕ is also a bijection. So we get the following theorem.

Theorem 12.10. *The group $(T, \circ)$ is isomorphic to the group $(\mathbb{C}, +)$.*

12.3.6 Exercises

1. Let T be the translation by the vector $\langle 1, 3 \rangle$.

   a. Write an algebraic representation for the translation from each of the four perspectives.
   b. Find the image of the circle $C = \{(x, y) \in \mathbb{R}^2 \mid (x-2)^2 + (y+3)^2 = 9\}$ under this translation.
   c. Let $A = (1, 2)$, $B = (-2, 3)$ and $C = (0, 0)$. Find the image of the triangle $\triangle ABC$ under this translation.

2. Let $z \in \mathbb{C}$. Why is $T_{z+\overline{z}}$ a horizontal translation?

3. Let $z \in \mathbb{C}$. Show that $T_{z-\overline{z}}$ is a vertical translation.

4. Show that the set of horizontal translations with function composition forms a group and that the group is isomorphic to $(\mathbb{R}, +)$.

5. We say that two translations, T_1 and T_2, are parallel if for any point A, A, $T_1(A)$, and $T_2(A)$ are co-linear. Assuming that T_1 and T_2 are parallel, answer the following questions.

   a. Let $T_1(x, y) = (x + h_1, y + k_1)$ and $T_2(x, y) = (x + h_2, y + k_2)$. Write a single algebraic equation that expresses the relationship between h_1, h_2, k_1, and k_2.
   b. Let $T_1(z) = z + b_1$ and $T_2(z) = z + b_2$. Write a single algebraic equation that expresses the relationship between b_1 and b_2. (Do not break the complex numbers up into their real and imaginary parts.)

6. Is $(T, \circ)$ a normal subgroup of $(I, \circ)$? (Consider $f^{-1} \circ T_d \circ f$ for all $f \in I$.) If so, describe the elements of the factor group and determine if the factor group is isomorphic to another known group.

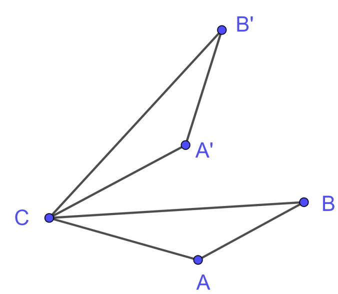
12.4 Rotations

If we represent the plane by a piece of paper, the rotations correspond to holding down a point on the paper and rotating the paper around that point. This means that as long as your rotation is not a multiple of 2π, the only fixed point of the plane is the center of the rotation. Intuitively we can see that this transformation of the plane does not change the paper is any way and so is an isometry of the plane.

12.4.1 Synthetic Plane

Definition 12.14. Given a point C and an angle θ, we define the rotation about C by θ for every point A in the plane by $A' = R_{(C,\theta)}(A)$ is the unique point in the plane so that $\overline{CA}$ is the same length as $\overline{CA'}$ and the measure of angle $\angle ACA'$ equals θ.

If we let C and α be given so that $R_{(C,\alpha)}$ is a rotation, for any two points A and B, we can let $A' = R_{(C,\alpha)}(A)$ and $B' = R_{(C,\alpha)}(B)$. If A, B, and C are not co-linear, we can show that triangles $\triangle CAB$ and $\triangle CA'B'$ have equal sides and angles.



From the definition of a rotation, we know that $\overline{CA}$ and $\overline{CA'}$ have the same length, and $\overline{CB}$ and $\overline{CB'}$ have the same length. Similarly angles $\angle ACA'$ and $\angle BCB'$ are equal. From this, we can see that angles $\angle ACB$ and $\angle A'CB'$ are equal. So by the SAS theorem, we have that $\overline{AB}$ and $\overline{A'B'}$ have the same length.

If A, B, and C are co-linear, one can use properties of line segments to prove that $\overline{AB}$ is equal to $\overline{A'B'}$.

12.4.2 Complex Plane

If $z = re^{i\alpha}$, we can see that $re^{i(\alpha+\theta)} = re^{i\alpha}e^{i\theta} = e^{i\theta}z$ is the rotation around the origin of z. We can then provide notation for these rotations as

$$R_{(0,\theta)}(z) = e^{i\theta} \cdot z.$$

If we have a point $c \in \mathbb{C}$ and an angle θ we can construct the rotation about c of angle θ through a composition of translations and the rotation about the origin. We begin the process by translating the point c to the origin. We can then rotate around the origin by θ and then translate the origin back to c. So the rotation about c by an angle θ is given by

$$T_c \circ R_{(0,\theta)} \circ T_{-c}.$$

Definition 12.15. Given an angle θ, we define the rotation about a complex number c as

$$R_{(c,\theta)}(z) = e^{i\theta} \cdot (z - c) + c.$$

Since

$$R_{(c,\theta)}(z) = e^{i\theta} \cdot (z - c) + c = e^{i\theta} z + \left(1 - e^{i\theta}\right) c,$$

we can see that $R_{(c,\theta)}(z)$ is a complex function of the form $f(z) = az + b$ with $|a| = 1$ and if the angle of rotation is not an integer multiple of 2π, $a \neq 1$. The following theorem shows that these functions are equivalent.

Theorem 12.11. *Let $f : \mathbb{C} \to \mathbb{C}$ be a function. Then f is a rotation if and only if $f(z) = az + b$ for some $a \in \mathbb{C} \setminus \{1\}$ with $|a| = 1$ and for some $b \in \mathbb{C}$, or for $a = 1$ and $b = 0$.*

Proof. If $f(z) = az + b$ with $|a| = 1$ we can discover information about f by looking at its fixed points, the points $c \in \mathbb{Z}$ such that $f(c) = c$. We can then see that the fixed point of f corresponds to the points where $ac + b = c$, or when

$$c = \frac{b}{1 - a}.$$

So if $a \neq 1$, there is a single fixed point and we can see that

$$f(z) = a\left(z - \frac{b}{1 - a}\right) + \frac{b}{1 - a}$$

and that f is a rotation. If $a = 1$ and $b = 0$, we see that we have the identity, which can be considered a rotation of angle 0 about any point.

If f is a rotation, then

$$f(z) = e^{i\theta}(z - c) + c = e^{i\theta} z + c - ce^{i\theta} = az + b,$$

where $a = e^{i\theta}$, so $|a| = 1$ and $b = c - ce^{i\theta}$. If $a = 1$, then $f(z) = z$ and so $b = 0$. $\square$

12.4.3 Cartesian Plane and Vector Space

Now that we know the representations of rotations from the perspective of the complex plane we can use that information to find the representations for rotations from the other perspectives.

If we let $z = x + iy$, we can see that

$$R_{(0,\theta)}(z) = (\cos(\theta) + i\sin(\theta))(x + iy) = (x\cos(\theta) - y\sin(\theta)) + i\left(x\sin(\theta) + y\cos(\theta)\right).$$

So in the Cartesian plane perspective, the rotation about the origin by angle θ can be written as

$$R_{((0,0),\theta)}\left((x, y)\right) = \left(x\cos(\theta) - y\sin(\theta), x\sin(\theta) + y\cos(\theta)\right).$$

In the perspective of the vector space we have

$$R_{\langle 0,0\rangle,\theta}\left(\langle x, y\rangle\right) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The algebraic representations for the general rotations can be found using similar methods as used in the complex plane perspective.

Related Content Standards

- (HSG.GPE.5) Prove the slope criteria for parallel and perpendicular lines and use them to solve geometric problems (e.g., find the equation of a line parallel or perpendicular to a given line that passes through a given point).

For the rotation by $\frac{\pi}{2}$ about a point (h,k) can then be written as

$$R_{((h,k),\frac{\pi}{2})}\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}0&-1\\1&0\end{pmatrix}\begin{pmatrix}x-h\\y-k\end{pmatrix}+\begin{pmatrix}h\\k\end{pmatrix}=\begin{pmatrix}-y+k+h\\x-h+k\end{pmatrix}$$

If we apply this rotation to the line $y=m(x-h)+k$, we have

$$R_{((h,k),\frac{\pi}{2})}\begin{pmatrix}x\\m(x-h)+k\end{pmatrix}=\begin{pmatrix}-(m(x-h)+k)+k+h\\x-h+k\end{pmatrix}=\begin{pmatrix}-m(x-h)+h\\x-h+k\end{pmatrix}$$

and since

$$\frac{-1}{m}\left((-m(x-h)+h)-h\right)+k=x-h+k$$

we see that the image of the line under the rotation is $y=\frac{-1}{m}(x-h)+k$. So the slope of perpendicular lines are negative multiplicative reciprocals of each other.

12.4.4 Compositions and Groups of Rotations

If we have two rotations centered at the same point c, $R_{(c,\theta)}$ and $R_{(c,\phi)}$, we see that the composition

$$R_{(c,\theta)}\circ R_{(c,\phi)}(z)=e^{i\theta}\left(\left(e^{i\phi}(z-c)+c\right)-c\right)+c=e^{i(\theta+\phi)}(z-c)+c$$

is a rotation centered at c with the new angle of rotation the sum of the original two angles,

$$R_{(c,\theta)}\circ R_{(c,\phi)}=R_{(c,\theta+\phi)}.$$

Hence with the same center, compositions of rotations is closed under composition. Since $R_{(c,0)}(z)=z$ and $R^{-1}_{(c,\theta)}=R_{(c,-\theta)}$ we see that the set of rotations about a single fixed point forms a group under function composition. For each point $c\in\mathbb{C}$, we can let R_c be the set of rotations centered at c.

Theorem 12.12. *For each complex number $c\in\mathbb{C}$, the set of rotations centered at c with the operation of composition, $(R_c,\circ)$, is a group.*

Since the addition of the angles is commutative, the order of operations of the compositions of these two rotations does not matter. Thus the group $(R_c,\circ)$ is abelian.

Since $R_{(c,\theta)}\circ R_{(c,\psi)}=R_{(c,\theta+\psi)}$, our first instinct is that the composition of rotations about a single center corresponds with the addition of real numbers. However, because of the periodicity of the trigonometric functions, for any angle θ, we see that $R_{(c,\theta)}=R_{(c,\theta+2k\pi)}$ for any integer k. We can instead let

$$S^1=\{z\in\mathbb{C}\mid |z|=1\}$$

be the unit circle in the complex plane. We also know that S^1 can be expressed using angles as

$$S^1=\left\{e^{i\theta}\mid\theta\in\mathbb{R}\right\}$$

with $e^{i\theta}\cdot e^{i\psi}=e^{i(\theta+\psi)}$. Therefore, $(S^1,\cdot)$ forms a group with identity $e^{i0}=1$ and inverses $\left(e^{i\theta}\right)^{-1}=e^{i(-\theta)}$. It is then straight forward to see that this group is isomorphic to $(R_c,\circ)$ with isomorphism $\phi:S^1\to R_c$ defined by

$$\phi\left(e^{i\theta}\right)=R_{(c,\theta)}.$$

Theorem 12.13. *For each complex number $c \in \mathbb{C}$, the group $(R_c, \circ)$ is isomorphic to the unit circle in $\mathbb{C}$, S^1, with the operation of multiplication, $(S^1, \cdot)$.*

If we look at the composition of two rotations about two different centers we have that

$$\begin{aligned} R_{(c,\theta)} \circ R_{(d,\psi)}(z) &= e^{i\theta}\left(e^{i\psi}(z-d)+d-c\right)+c \\ &= e^{i(\theta+\psi)}z + \left(e^{i\theta}\left(1-e^{i\psi}\right)d + (1-e^{i\theta})c\right) \end{aligned}$$

If $\theta+\psi$ is an integer multiple of 2π, we see that $e^{i(\theta+\psi)} = 1$. So

$$R_{(c,\theta)} \circ R_{(d,\psi)}(z) = z + (e^{i\theta}-1)d + (1-e^{i\theta})c = z + (e^{i\theta}-1)(d-c).$$

Therefore, the composition of the two rotations is a translation. We note that this means that the set of all rotations is not closed under function composition and so does not form a group.

If $\theta+\psi$ is not an integer multiple of 2π we can see that the composition of the two rotations is a function of the form $f(z) = az+b$ with $a = e^{i(\theta+\psi)}$. The earlier work on rotations shows that this is a rotation centered at

$$\frac{b}{1-a} = \frac{e^{i\theta}\left(1-e^{i\psi}\right)d + (1-e^{i\theta})c}{1-e^{i(\theta+\psi)}}.$$

Combining the translations and rotations into a single set, we see that we have another subgroup of I.

Definition 12.16. Let P be the set of orientation preserving isometries of the plane.

$$P = \{f : \mathbb{C} \to \mathbb{C} \mid f(z) = az+b \text{ for some } a, b \in \mathbb{C} \text{ with } |a| = 1\}$$

Theorem 12.14. *The set of orientation preserving isometries, P, is a subgroup of the isometries, I.*

12.4.5 Exercises

1. Express the rotation $R_{(A,\frac{\pi}{2})}$, where $A = (0,1)$, as a transformation in each of the different perspectives.
2. Let R be the rotation about the point $(1,1)$ with an angle of $\frac{\pi}{3}$.
   a. Write an algebraic representation for the rotation from each of the four perspectives.
   b. Find the image of the circle $C = \left\{(x,y) \in \mathbb{R}^2 \mid (x-2)^2 + (y+3)^2 = 9\right\}$ under this rotation.
   c. Let $A = (1,2)$, $B = (-2,3)$ and $C = (0,0)$. Find the image of the triangle $\triangle ABC$ under this rotation.
3. Prove that the set of all rotations of the plane with the operation of composition is not a group.
4. Show that the order of composition matters when composing a rotation and a translation.
5. Let f and g be two rotations such that $g \circ f$ is a translation.
   a. If $f(z) = a(z-c)+c$ and $g(z) = b(z-d)+d$, what relationships exist between the coefficients of f and g?
   b. What relationships exist between $f \circ g$ and $g \circ f$?
6. Consider the group of orientation preserving isometries, P, in relation to the group of isometries, I. Is P a normal subgroup of I? If so, describe the elements of the factor group and determine if the factor group is isomorphic to another known group.
7. Let $c \in \mathbb{C}$ be a fixed point on the plane. Since the set of rotations about c form a group, R_c is a subgroup of I and P.

a. Is R_c a normal subgroup of I? If so, describe the elements of the factor group and determine if the factor group is isomorphic to another known group.
b. Is R_c a normal subgroup of P? If so, describe the elements of the factor group and determine if the factor group is isomorphic to another known group.

8. Is T, the subgroup of translations a normal subgroup of P? If so, describe the elements of the factor group and determine if the factor group is isomorphic to another known group.

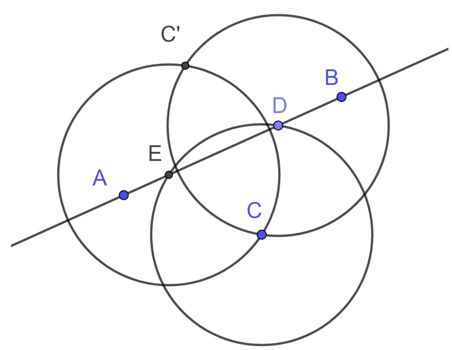
12.5 Reflections

Continuing our illustrations of isometries using a piece of paper to represent the plane, a reflection corresponds to flipping the paper over while keeping one line of the paper fixed. For example, holding two opposite corners of the paper still while flipping over the diagonal line.

12.5.1 Synthetic Plane

Theorem 12.15. *Given a line $\overleftrightarrow{AB}$ and a point C not on $\overleftrightarrow{AB}$, there exists a unique point C' such that $\overline{CC'}$ is perpendicular to $\overleftrightarrow{AB}$ and $\overleftrightarrow{AB}$ intersects $\overline{CC'}$ at its midpoint. We will call the point C' the reflection of C about $\overleftrightarrow{AB}$.*

As in the figure below, for a given line $\overleftrightarrow{AB}$ and a point C not on the line we can find points D and E on $\overleftrightarrow{AB}$ that are equidistant from C. We then construct the circles centered at D and E through the point C. These circles then intersect at another point, C', that is the reflection of C about $\overleftrightarrow{AB}$.



Definition 12.17. Given a line $\overleftrightarrow{AB}$ we define the reflection about $\overleftrightarrow{AB}$, $r_{\overleftrightarrow{AB}}$, for a point C as C if $C \in \overleftrightarrow{AB}$ and as its reflection about $\overleftrightarrow{AB}$ otherwise.

12.5.2 Complex Plane

We see that the function $f(z) = \overline{z}$ is a reflection about the real axis. In order to define a reflection about a generic line we can first use translations and rotations to map that line to the real axis, perform the reflection, and then transform the real axis back to the original line. So if l is a line, we can choose a point c on the line and we can let θ be the angle that the line makes with the horizontal axis. This point and angle then define the line,

$$l = \left\{z \in \mathbb{C} \mid z = c + re^{i\theta} \text{ for some } r \in \mathbb{R}\right\}.$$

So by translating the plane so that c maps to 0, T_{-c}; rotating the plane so that the image of the line maps to the real axis, $R_{(0,-\theta)}$; reflecting about the real axis, $r_{\mathrm{Im}(z)=0}$; rotating the real axis back to a line with angle θ with the horizontal, $R_{(0,\theta)}$; and finally a translation that maps 0 back to c, T_c, we have the reflection about the line l,

$$r_l(z) = T_c \circ R_{(0,\theta)} \circ r_{\mathrm{Im}(z)=0} \circ R_{(0,-\theta)} \circ T_{-c}(z) = e^{i(2\theta)}\,\overline{(z-c)} + c.$$

Definition 12.18. Let l be a line through a point c making an angle θ with the real axis. Then we define

$$r_{(c,\theta)}(z) = e^{i(2\theta)}\,\overline{(z-c)} + c$$

and see that this function is a reflection about the line l.

We can see from the expression for the reflections that they are of the form $f(z) = a\overline{z} + b$ for complex numbers a and b, with $|a| = 1$. However, we see in the following theorem that not all functions of this form are reflections.

Theorem 12.16. *Let $f(z) = a\overline{z} + b$ be a function such that $|a| = 1$. Then f is a reflection if and only if $a\overline{b} + b = 0$.*

Proof. If we first assume that f is a reflection, then there is a point c and an angle θ such that

$$f(z) = e^{i(2\theta)}\,\overline{(z-c)} + c.$$

So $f(z) = a\overline{z} + b$ for $a = e^{i(2\theta)}$, so that $|a| = 1$, and $b = c - e^{i(2\theta)}\,\overline{c}$. Then

$$a\overline{b} + b = e^{i(2\theta)} \cdot \left(\overline{c} - e^{-i(2\theta)}\,c\right) + c - e^{i(2\theta)}\,\overline{c} = 0.$$

If we let $f(z) = a\overline{z} + b$ for some $a, b \in \mathbb{C}$ such that $|a| = 1$ and $a\overline{b} + b = 0$, there exists an angle θ such that $a = e^{i(2\theta)}$. We also know that

$$a\overline{z} + b = a\overline{z} + \frac{b}{2} - \frac{a\overline{b}}{2} + \frac{b}{2} + \frac{a\overline{b}}{2} = a\overline{z} - a\frac{\overline{b}}{2} + \frac{b}{2} + \frac{a\overline{b}+b}{2} = a\overline{\left(z - \frac{b}{2}\right)} + \frac{b}{2} + \frac{a\overline{b}+b}{2}.$$

Since $a\overline{b} + b = 0$, we see that

$$f(z) = e^{i(2\theta)}\,\overline{\left(z - \frac{b}{2}\right)} + \frac{b}{2}$$

and so f is a reflection about the line through $\frac{b}{2}$ making an angle θ with the real axis. □

12.5.3 Vector Space and Cartesian Plane

As we did with rotations, we will use the information from the Complex Plane representation to determine the representation of reflections in the Vector Space and Cartesian Plane representations.

Reflections over the horizontal axis are described by $\langle x, y\rangle \mapsto \langle x, -y\rangle$. So using matrices, this reflection can be represented by the matrix

$$r_{y=0} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

If we combine this matrix with the transformations of rotation and translation we see that the reflection about a line, l, through the point (h, k) making an angle θ with the horizontal is described as

$$\begin{aligned} r_l(\langle x, y\rangle) &= \begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ -\sin(2\theta) & \cos(2\theta) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x-h \\ y-k \end{pmatrix} + \begin{pmatrix} h \\ k \end{pmatrix} \\ &= \begin{pmatrix} (x-h)\cos(2\theta) - (y-k)\sin(2\theta) + h \\ -(x-h)\sin(2\theta) - (y-k)\cos(2\theta) + k \end{pmatrix}. \end{aligned}$$

So the Cartesian Plane representation of a reflection about a line l through a point (h, k) making an angle θ with the horizontal is given by

$$r_l(x, y) = ((x-h)\cos(2\theta) - (y-k)\sin(2\theta) + h, -(x-h)\sin(2\theta) - (y-k)\cos(2\theta) + k)\,.$$

For the special case of a reflection about the vertical axis, we let $(h, k) = (0, 0)$ and $\theta = \frac{\pi}{2}$ to see that $r_{x=0}(x, y) = (-x, y)$. For the reflection about the line $y = x$, we can use the points $(0, 0)$ and the angle $\frac{\pi}{4}$ to see that $r_{y=x}(x, y) = (y, x)$.

12.5.4 Compositions

If $r_1(z) = e^{i(2\theta_1)}\,\overline{(z - c_1)} + c_1$ and $r_2(z) = e^{i(2\theta_2)}\,\overline{(z - c_2)} + c_2$ are two generic reflections, we see that the composition of the two reflections is

$$\begin{aligned} r_1 \circ r_2(z) &= e^{i(2\theta_1)}\,\overline{\left(e^{i(2\theta_2)}\,\overline{(z - c_2)} + c_2 - c_1\right)} + c_1 \\ &= e^{i(2\theta_1)}\,\left(e^{-i(2\theta_2)}(z - c_2) + \overline{c_2 - c_1}\right) + c_1 \\ &= e^{i2(\theta_1 - \theta_2)}z + c_3 \end{aligned}$$

where $c_3 = -e^{i2(\theta_1 - \theta_2)}c_2 + e^{i(2\theta_1)}\,\overline{(c_2 - c_1)} + c_1$.

If the two reflections are the same, we see that the composition generates the identity function.

Theorem 12.17. *If m is a line in the plane with r_m denoting the reflection about m, then $r_m^{-1} = r_m$. In other words, each reflection is its own inverse.*

If the two lines are distinct, $r_1 \circ r_2$ is a transformation that is a translation if $e^{i2(\theta_1 - \theta_2)} = 1$, and a rotation otherwise.

When the composition is a translation, the difference between the angles of the two lines of reflection is a multiple of π and so the two lines are parallel. So we can assume that $\theta_1 = \theta_2$ and so

$$\begin{aligned} r_1 \circ r_2(z) &= z - c_2 + e^{i(2\theta_1)}\overline{c_2} - e^{i(2\theta_1)}\overline{c_1} + c_1 \\ &= z + e^{i\theta_1}\left(e^{i\theta_1}\overline{c_2} - e^{-i\theta_1}c_2 + e^{-i\theta_1}c_1 - e^{i\theta_1}\overline{c_1}\right) \\ &= z + e^{i\theta_1}\left(e^{i\theta_1}\overline{(c_2 - c_1)} - e^{-i\theta_1}(c_2 - c_1)\right) \\ &= z + e^{i\theta_1}\left(2i\mathrm{Im}\left(e^{i\theta_1}\overline{(c_2 - c_1)}\right)\right) \\ &= z + e^{i(\theta_1 + \frac{\pi}{2})}2\mathrm{Im}\left(e^{i\theta_1}\overline{(c_2 - c_1)}\right), \end{aligned}$$

which is a translation in the direction perpendicular to the two lines, with a magnitude of twice the distance between the two lines, and in the direction from the line of reflection of r_2 towards the line of reflection of r_1.

When the composition is a rotation, we know that the lines are not parallel and so they intersect. Without any loss of generality we can rewrite the equations of the two reflections through this point c as

$$r_1(z) = e^{i(2\theta_1)}\,\overline{(z - c)} + c \quad \text{and} \quad r_2(z) = e^{i(2\theta_2)}\,\overline{(z - c)} + c$$

and the composition as

$$r_1 \circ r_2(z) = e^{i2(\theta_1 - \theta_2)}\,\overline{(z - c)} + c.$$

So we see that the center of rotation is the intersection point of the two lines. The angle of rotation is then twice the angle from the line of reflection for r_2 towards the line of reflection of r_1.

We combine all of this information into the following theorem.

Theorem 12.18 (Two-Reflection Theorem)**.** *Let m and n be two distinct lines in the plane and let r_m and r_n be the reflections about the lines.*

If m and n intersect at a point c, then $r_n \circ r_m$ is a rotation centered at c with angle of rotation double the angle from m to n. Furthermore, $r_m \circ r_n$ is a rotation centered at c with angle of rotation double the angle from n to m.

If m and n are parallel lines, $r_n \circ r_m$ is a translation that is perpendicular to the two lines with a magnitude twice the distance between the lines in the direction from m to n. Similarly, $r_m \circ r_n$ is a translation perpendicular to the lines with magnitude of twice the distance from n to m.

12.5.5 Exercises

1. Use properties of triangles and quadrilaterals to show that the reflections defined on the synthetic plane are isometries. (Make sure to include all possible cases of the relationships between the two points and the line of reflection.)
2. Let r be the reflection across the line $y = \sqrt{3}x$.
   a. Write an algebraic representation for the reflection from each of the four Cartesian, Vector, and Complex perspectives.
   b. Find the image of the circle $C = \left\{(x, y) \in \mathbb{R}^2 \mid (x-2)^2 + (y+3)^2 = 9\right\}$ under this reflection.
   c. Let $A = (1, 2)$, $B = (-2, 3)$ and $C = (0, 0)$. Find the image of the triangle $\triangle ABC$ under this reflection.
3. For the set of reflections, identify whether each the properties of abelian groups hold. For each property, prove it holds or identify an appropriate argument or counter-example as to why it does not.
   a. closure
   b. identity
   c. inverse
   d. commutativity

12.6 Glide-Reflections

Since reflections and translations are both isometries, the composition of a reflections and a translation is also an isometry.

12.6.1 Synthetic Plane

Definition 12.19. Let A and B be distinct points in the plane. The function

$$g_{\overrightarrow{AB}} = r_{\overleftrightarrow{AB}} \circ T_{\overrightarrow{AB}}$$

is called the **glide-reflection** of $\overrightarrow{AB}$.

Since the translation is parallel to the line of reflection, the order of the reflection and translation are interchangeable (the proof is left as an exercise). This means that

$$\begin{aligned} g_{\overrightarrow{AB}} \circ g_{\overrightarrow{BA}} &= \left(r_{\overleftrightarrow{AB}} \circ T_{\overrightarrow{AB}}\right) \circ \left(r_{\overleftrightarrow{AB}} \circ T_{\overrightarrow{BA}}\right) \\ &= r_{\overleftrightarrow{AB}} \circ \left(T_{\overrightarrow{AB}} \circ T_{\overrightarrow{BA}}\right) \circ r_{\overleftrightarrow{AB}} \\ &= r_{\overleftrightarrow{AB}} \circ r_{\overleftrightarrow{AB}} = \text{identity} \end{aligned}$$

and so $g^{-1}_{\overrightarrow{AB}} = g_{\overrightarrow{BA}}$.

Theorem 12.19. *Every glide-reflection is a composition of three reflections.*

Proof. Let A and B be distinct points in the plane and let C be the midpoint of $\overline{AB}$. Let m and n be lines through A and C, respectively, such that m and n are both perpendicular to $\overleftrightarrow{AB}$. The two-reflection theorem implies that $T_{\overrightarrow{AB}} = r_n \circ r_m$ and so

$$g_{\overrightarrow{AB}} = r_{\overleftrightarrow{AB}} \circ r_n \circ r_m$$

and so every glide-reflection is a composition of three reflections. □

12.6.2 Complex Plane

We know from Theorem 12.7 that every isometry $f : \mathbb{C} \to \mathbb{C}$ is of the form

$$f(z) = az + b \quad \text{or} \quad f(z) = a\overline{z} + b$$

for some $a, b \in \mathbb{C}$ with $|a| = 1$. We have shown that isometries of the form $f(z) = az+b$ are either translations or rotations and isometries of the form $f(z) = a\overline{z} + b$ are reflections if $a\overline{b} + b = 0$. The following theorem shows that the glide-reflections complete our types of isometries.

Theorem 12.20. *Let $f(z) = a\overline{z} + b$ be a function such that $|a| = 1$. Then f is a glide-reflection with a non-zero glide if and only if $a\overline{b} + b \neq 0$.*

Proof. In the proof of Theorem 12.16 we showed that for complex numbers a and b, with $|a| = 1$, that

$$f(z) = a\overline{z} + b = a\overline{\left(z - \frac{b}{2}\right)} + \frac{b}{2} + \frac{a\overline{b} + b}{2}.$$

Since $|a| = 1$, $a = e^{i2\theta}$ for some angle θ.

We can then let m be the line through $\frac{b}{2}$ making an angle of θ with the horizontal.

Furthermore,

$$\frac{a\overline{b} + b}{2} = \frac{e^{i2\theta}\overline{b} + b}{2} = e^{i\theta}\frac{e^{i\theta}\overline{b} + e^{-i\theta}b}{2} = e^{i\theta} \cdot \mathrm{Re}\left(e^{i}\theta\overline{b}\right).$$

Therefore, this complex number is in the same direction as m and we see that

$$f(z) = a\overline{z} + b = r_m \circ T_{\frac{a\overline{b}+b}{2}}$$

is a glide-reflection with a non-zero glide. □

Let

$$G = \{f : \mathbb{C} \to \mathbb{C} |\, f(z) = a\overline{z} + b \text{ for some } a, b \in \mathbb{C} \text{ with } |a| = 1\}$$

be the set of glide reflections. Then we can see that the set of isometries, I, is the union of the set of orientation preserving isometries, P, and G. So all orientation reversing isometries are glide reflections, with or without the glide.

12.6.3 Exercises

1. Prove that for $g_{\overrightarrow{AB}}$,
$$r_{\overleftrightarrow{AB}} \circ T_{\overrightarrow{AB}} = T_{\overrightarrow{AB}} \circ r_{\overleftrightarrow{AB}}.$$

2. Let g be the glide-reflection about the vector $\langle 1, 1\rangle$.

a. Write an algebraic representation for the glide-reflection from the Cartesian, Vector, and Complex perspectives.
b. Find the image of the circle $C = \left\{(x, y) \in \mathbb{R}^2 \mid (x-2)^2 + (y+3)^2 = 9\right\}$ under this glide-reflection.
c. Let $A = (1, 2)$, $B = (-2, 3)$ and $C = (0, 0)$. Find the image of the triangle $\triangle ABC$ under this glide-reflection.

3. Let A and B be any two distinct points. Prove that the composition

$$R_{B,\pi} \circ r_{\overleftrightarrow{AB}} \circ R_{A,\pi}$$

is a glide-reflection and find its axis and the vector of translation.

4. Is G a normal subgroup of I? If so, describe the elements of the factor group and determine if the factor group is isomorphic to another known group.

12.7 Group of Isometries

Now that we know that every isometry is a translation, rotation, reflection, or glide-reflection, we will look at this set of isometries more deeply and see how they all fit together.

Related Content Standards

- (HSG.CO.5) Given a geometric figure and a rotation, reflection, or translation, draw the transformed figure using, e.g., graph paper, tracing paper, or geometry software. Specify a sequence of transformations that will carry a given figure onto another.
- (HSG.CO.6) Use geometric descriptions of rigid motions to transform figures and to predict the effect of a given rigid motion on a given figure; given two figures, use the definition of congruence in terms of rigid motions to decide if they are congruent.

We will call the translations and rotations **orientation preserving** because they preserve the order of the vertices of a triangle when viewed in a clockwise direction. The reflections and glide-reflections, on the other hand, change the order of the vertices and so we call these isometries **orientation reversing**.

Theorem 12.21 (Three-Reflection Theorem)**.** *Every isometry can be written as a composition of at most three reflections.*

Proof. Let f be an isometry. Then f is a translation, rotation, reflection, or glide-reflection.

If f is a translation that takes a point A to a point B, then we can find the midpoint of $\overline{AB}$ and call that point C. We can then construct lines m and n through A and C, respectively, so that m and n are perpendicular to $\overleftrightarrow{AB}$. Then $f = r_n \circ r_m$ and so can be written as a composition of two reflections.

If f is a rotation through the point $C = (h, k)$ with an angle of θ, we can use the Cartesian perspective and let m be the line $y = k$ and let n be the line $y = \tan\left(\frac{\theta}{2}\right)(x-h)+k$. Then $f = r_n \circ r_m$ and so can be written as a composition of two reflections.

Any reflection is already written as a composition of a reflection.

Since glide-reflections are a composition of a reflection and a translation, and since translations can be written as a composition of two reflections, glide-reflections can be written as a composition of three reflections. $\square$

From the proof above, we see that transformations that can be written as an odd number of reflections are orientation reversing, while those that can be written as an even number of reflections are orientation preserving. So, in a way, the orientation of the isometry is similar to a number being positive or negative.

When looking at the complex plane representation of the isometries, we see that the orientation preserving isometries are of the form $f(z) = az+b$ and the orientation reversing isometries are of the form $f(z) = a\overline{z}+b$. We will use this information to see how the behavior of an isometry on three distinct points determines the isometry.

Theorem 12.22 (Three Point Theorem)**.** *Let A, B, and C be three points in the plane that are not co-linear. Then any isometry, f, is determined by the locations of $f(A)$, $f(B)$, and $f(C)$.*

Proof. We will let f be an isometry. We will use the perspective of the complex plane and let z_1, z_2, and z_3 be three points that are not co-linear. Then we set $w_1 = f(z_1)$, $w_2 = f(z_2)$, and $w_3 = f(z_3)$.

In order to determine the orientation of the isometry we will use properties of the angles between the segments connecting the corresponding complex numbers. We can set z_1 and w_1 as our points of reference and then look at the relationship of the other points to these. We can measure the angle from $z_2 - z_1$ to $z_3 - z_1$ as

$$\arg(z_3 - z_1) - \arg(z_2 - z_1) = \arg\left(\frac{z_3 - z_1}{z_2 - z_1}\right).$$

Similarly, the angle from $w_2 - w_1$ to $w_3 - w_1$ is

$$\arg\left(\frac{w_3 - w_1}{w_2 - w_1}\right).$$

If these arguments are the same, then the isometry is orientation preserving and if they are additive inverses of each other the isometry is orientation reversing. This can most easily be checked by comparing the ratios

$$\frac{w_3 - w_1}{w_2 - w_1} \quad \text{and} \quad \frac{z_3 - z_1}{z_2 - z_1}.$$

If the ratios are equal, then the isometry is orientation preserving. If the ratios are complex conjugates of each other the isometry is orientation reversing.

Once we have determined if the isometry is orientation preserving or orientation reversing we can see how the equation of the isometry is determined.

If f is orientation preserving, $f(z) = az + b$ for some $a, b \in \mathbb{C}$ with $|a| = 1$. Using the three points we have the following system of equations

$$\begin{aligned} w_1 &= az_1 + b \\ w_2 &= az_2 + b \\ w_3 &= az_3 + b \end{aligned}$$

Using the first two equations we can determine that

$$a = \frac{w_2 - w_1}{z_2 - z_1}.$$

Then from the third point we see that

$$f(z) = \frac{w_2 - w_1}{z_2 - z_1}(z - z_3) + w_3.$$

If f is orientation reversing, $f(z) = a\overline{z} + b$ for some $a, b \in \mathbb{C}$ with $|a| = 1$. Then following a similar process to the one above we see that

$$f(z) = \frac{w_2 - w_1}{\overline{(z_2 - z_1)}}\overline{(z - z_3)} + w_3.$$

$\square$

12.7.1 Fixed Points

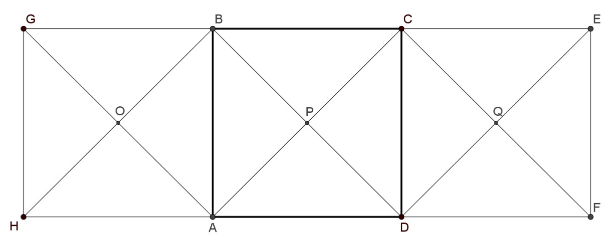
For a function $f : A \to A$, we say that $c \in A$ is a fixed point of f if $f(c) = c$. This property of fixed points is a way of describing different functions and can help us classify functions. In particular, the isometries of the plane have different numbers of fixed points.

12.7.2 Exercises

1. Classify the transformation f as one of the four types of isometries from its formula in the complex plane.

   a. $f(z) = \left(\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}\right) z + i$
   b. $f(z) = \left(\frac{1}{\sqrt{2}} - \frac{i}{\sqrt{2}}\right) \overline{z} + i$
   c. $f(z) = i\overline{z} - 2$
   d. $f(z) = z - 2 + i$

2. Identify the following compositions where $ABCD$ is the square below. (You may need to draw additional points.) \begin{figure}

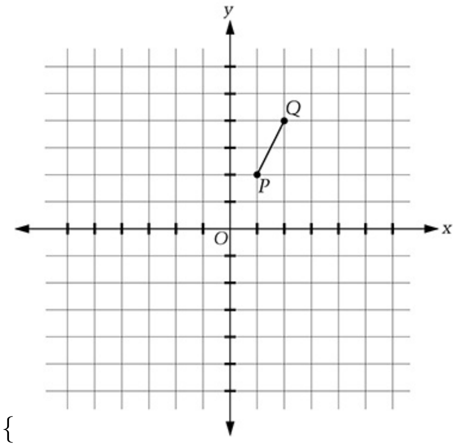
{ 

}

\end{figure}

```
a. ${\displaystyle R_{A,\frac{\pi}{2}} \circ R_{B,\frac{\pi}{2}}}$
a. ${\displaystyle R_{A,\frac{\pi}{2}} \circ T_{\stackrel{\longrightarrow}{CA}}}$
a. ${\displaystyle T_{\stackrel{\longrightarrow}{BC}} \circ T_{\stackrel{\longrightarrow}{AD}}}$
a. ${\displaystyle T_{\stackrel{\longrightarrow}{AB}} \circ R_{A,\frac{\pi}{3}}}$
a. ${\displaystyle T_{\stackrel{\longrightarrow}{AB}} \circ g_{\stackrel{\longrightarrow}{DC}}}$
a. ${\displaystyle r_{AD} \circ R_{C,\frac{\pi}{2}}}$
a. ${\displaystyle g_{\stackrel{\longrightarrow}{AC}} \circ g_{\stackrel{\longrightarrow}{BD}}}$
a. ${\displaystyle r_{AD} \circ T_{\stackrel{\longrightarrow}{AB}}}$
```

1. Perform the following two transformations on the graph below with each square representing 1 square unit. After each transformation, draw the resulting image of segment $\overline{PQ}$. \begin{figure}



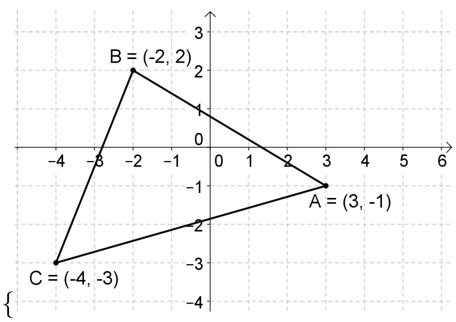
{

}

\end{figure}

```
a. Rotate the segment $90^\circ$ counterclockwise about point $P$ . Label the resulting segment $I$. Fi
a. Reflect the segment $I$ you drew in the first part across the $x$ -axis. Label the resulting segment
a.  Find an algebraic representation of the composition of the two transformations using a function of
```

1. What are the coordinates of the image of $\triangle ABC$ after a reflection across the y-axis followed by a rotation of 90° clockwise about the origin? Show your work, including a graph of the image of the triangle after each transformation, and explain how you found your answer. \begin{figure}

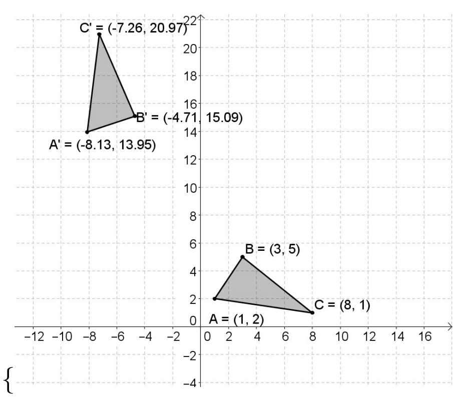


{

}

\end{figure}

1. Find an algebraic representation of the rigid motion that takes $\triangle ABC$ onto $\triangle A'B'C'$. Explain your method for finding the transformation and why that method would work. \begin{figure}



{

}

\end{figure}

1. Let I be the set of isometries of the plane, i.e.

$$I = \{f : \mathbb{C} \to \mathbb{C} | f(z) = az + b \text{ or } f(z) = a\bar{z} + b \text{ with } |a| = 1\}.$$

   Let $g \in I$, and $\phi_g : I \to I$ be the function defined by

$$\phi_g(f) = g^{-1} f g.$$

   Prove that ϕ_g

   a. maps the identity function to itself
   b. is one-to-one
   c. is onto
   d. maps translations to translations
   e. maps rotations to rotations
   f. maps reflections to reflections

2. Find all fixed points, if any, of the following plane transformations:

   a. The rotation $R_{C,\theta}$
   b. The translation T_b by the vector b
   c. The reflection r_m

3. For the given orientation preserving isometries, find two lines of reflection that will produce the described transformation. Express lines in the form $y = mx + b$ or $x = a$. Indicate the order they should be completed in.

   a. A rotation of $\frac{\pi}{3}$ radians in the counterclockwise direction around the point $(0, 0)$.
   b. A rotation of π radians in the clockwise direction around the point $(1, 2)$.
   c. A translation equivalent to the vector $\langle 3, 4\rangle$.

12.8 Dilations and Similarity

Related Content Standards

- (7.G.1) Solve problems involving scale drawings of geometric figures, including computing actual lengths and areas from a scale drawing and reproducing a scale drawing at a different scale.
- (8.G.3) Describe the effect of dilations, translations, rotations, and reflections on two-dimensional figures using coordinates.
- (8.G.4) Understand that a two-dimensional figure is similar to another if the second can be obtained from the first by a sequence of rotations, reflections, translations, and dilations; given two similar two-dimensional figures, describe a sequence that exhibits the similarity between them.
- (HSG.SRT.2) Given two figures, use the definition of similarity in terms of similarity transformations to decide if they are similar; explain using similarity transformations the meaning of similarity for triangles as the equality of all corresponding pairs of angles and the proportionality of all corresponding pairs of sides.

12.8.1 Synthetic Plane

12.8.2 Cartesian Plane

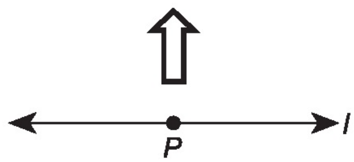
12.8.3 Vector Space

12.8.4 Complex Plane

12.8.5 Compositions

12.8.6 Exercises

1. Shane is working with transformations of an arrow shape about line l and point P. \begin{figure}



{

}

\end{figure}

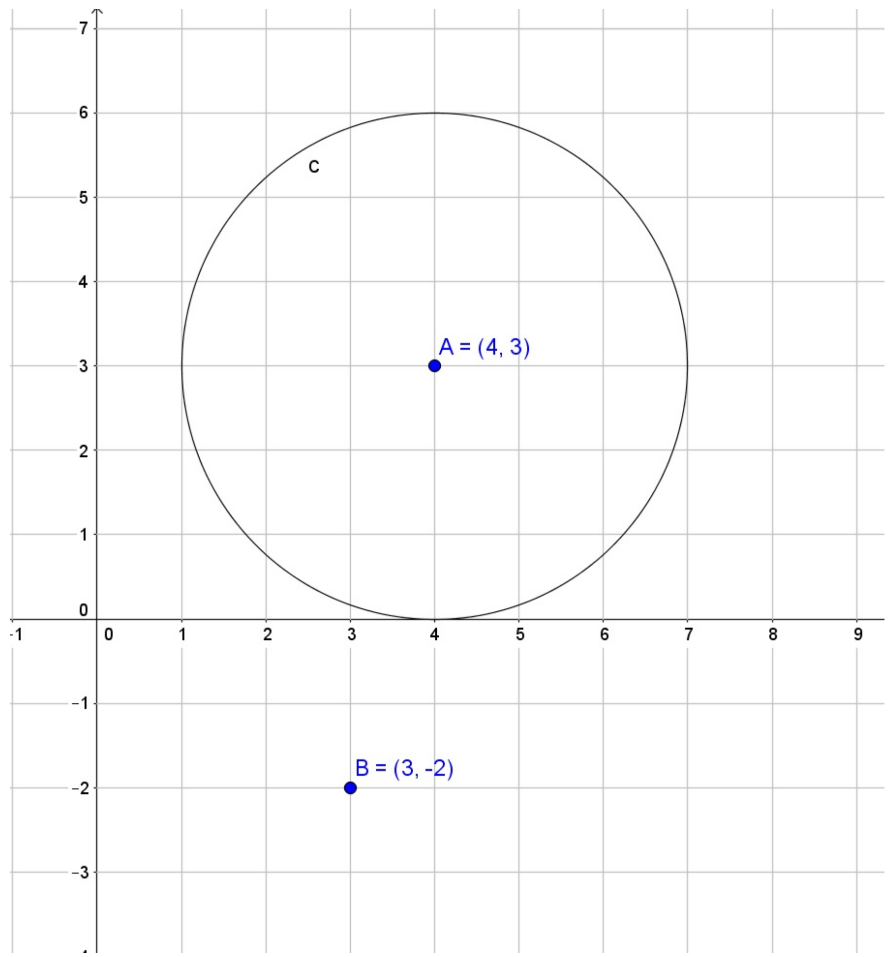
```
a. Draw the image of the arrow after 2 successive transformations: a reflection across $l$ and then a r

a. Draw a dilation of the original arrow centered at $P$ with a scale factor of 3.  Show your work, and
a. How does the area of the arrow after the dilation in Part b) compare to the area of the original arr
```

1. For each function below, identify whether $f(z)$ is a size transformation or isometry.

   i. If it is an isometry, identify what type of isometry it is and describe as much detail as possible (e.g., if it is a rotation, give the center and angle of rotation).
   ii. If it is a similarity transformation, identify the scalar of the size transformation.

   a. $f(z) = -3z + 2$
   b. $f(z) = (2 + 2\sqrt{3}i)z - (3 + i)$
   c. $f(z) = (-1 + \sqrt{3}i)\overline{z} + (2 - i)$
   d. $f(z) = \left(-\frac{1}{3}i\right) z$
   e. $f(z) = \left(\frac{3}{5} + \frac{4}{5}i\right) \overline{z} + 8 - 16i$
   f. $f(z) = z + 2 - 2i$
   g. $f(z) = \sqrt{3}\overline{z} - i$ a $f(z) = \left(-\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)(z - 2 + 3i) + 2 - 3i$
   h. $f(z) = i\overline{z} + (-2 + 2i)$
   i. $f(z) = \left(\frac{1}{2} + i\right) z$
   j. The composition of three reflections over three distinct lines intersecting at a single point.
   k. The composition of two reflections over two distinct intersecting lines.

2. Draw the image of the circle, c, and find the equation of the image, after a dilation centered at B with ratio of similitude of 0.5.



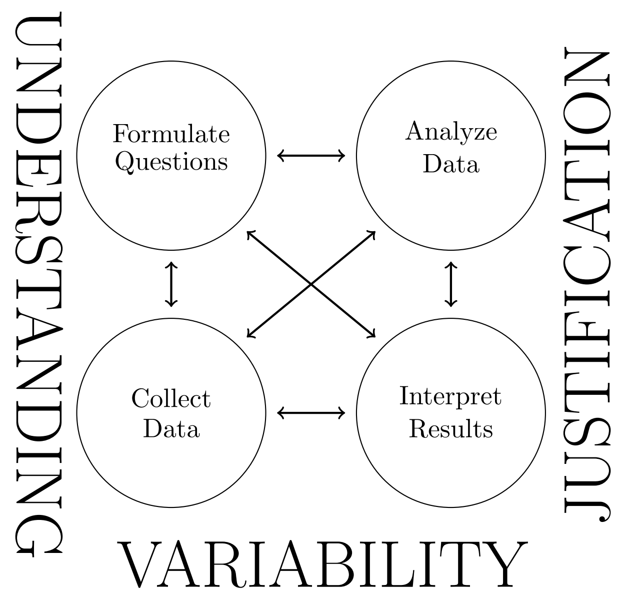
c
A = (4, 3)
B = (3, -2)

Part IV

DATA ANALYSIS

Chapter 13

Data Analysis Foundations



13.1 Statistics, Data Analysis, and Mathematics

Most people think of statistics and data analysis as part of the field of mathematics. While statistics and data analysis use a great many mathematical ideas, they are different in many ways. The development of statistics and data analysis out of mathematics in the nineteenth and twentieth centuries parallels the development of physics in the seventeenth and eighteenth centuries and computer science in the nineteenth and twentieth centuries. Each of these disciplines include vast applications of mathematics (physics uses fields such as calculus, differential geometry, and Hilbert spaces; computer science often uses set theory, abstract algebra, and numerical analysis; and statistics and data analysis makes use of topics such as probability, real analysis, and linear algebra). However, the language, goals, methods, and culture of these disciplines are distinct from one another.

Since the middle of the nineteenth century, the field of mathematics moved from the study of quantities to the study of abstract structures, built upon the foundation of logic and set theory. On the other hand, statistics as a discipline is grounded in the study of quantities within their original context. George Cobb and David Moore [1997] describe the differences very well.

> Although mathematicians often rely on applied context both for motivation and as a source of problems for research, the ultimate focus in mathematical thinking is on abstract patterns: the context is part of the irrelevant detail that must be boiled off over the flame of abstraction in order to reveal the previously hidden crystal of pure structure. In mathematics, context obscures structure. Like mathematicians, data analysts also look for patterns, but ultimately, in data analysis, whether the patterns have meaning, and whether they have any value, depends on how the threads of those patterns interweave with the complementary threads of the story line. In data analysis, context provides meaning. (p. 803)

In many ways one can think of **data analysis** as a process that involves the study of quantitative patterns within a situated context. Just as cooking takes many different raw ingredients and puts them together in unique ways using many different tools in order to create something to eat, data analysis takes raw data from a certain context and cleans it, transforms it, analyzes it, and reconfigures the generated information for consumption of better understanding of the context and making decisions about it. Hence, data analysis is an inherently inter-disciplinary process focused activity.

A **statistic** is a quantity computed from values in a collection of values. Some examples are a mean, median, or standard deviation computed from a set of numbers. It could also be a percentage of people that have a driver's license, or the number of people in a building at a certain time of day. It could also be the probability that someone will be diagnosed with cancer based upon other measurable health, physical, or sociological variables. The science of **statistics** uses these various quantities and ideas from the theory of probability distributions to process data and represent it in different ways. So we can think of statistics as a methodological discipline comprised of a vast set of tools used to analyze and interpret data.

A **descriptive statistic** is a summary statistic to quantitatively describe a collection of information. **Descriptive statistics** is the process of using those statistics to understand data and the context from which the data is derived. This could include describing properties of the data using graphs or tables. It could also describe how two variables are related using scatter plots and correlation coefficients. A key aspect of descriptive statistics is that these describe the data. These do not, by themselves, tell us anything about a population.

Inferential statistics uses samples to infer properties about a larger population, while descriptive statistics focuses on properties of the observed data. As such, inferential statistics uses descriptive statistics of a population of inference, **sample statistic**, along with assumptions about the underlying probability distribution of the population of interest to estimate a **population parameter**.

Example 13.1. Assume we are interested in exploring how height varies between boys and girls in the United States as they age. We ask our fellow teachers to collect data on age (in months) and height (in inches) for us, so we end up with data from 429 students between the ages of 11 and 17 from a single school district in the United States.

- (Population versus sample:) The population is the universe of things that fit the criteria of the thing you want to study. It is also described as the **population of interest or inference**. The sample is the set of objects on which you actually have measurements. In our example, the population of interest is the set of students in the United States. Our sample is the students between the ages of 11 and 17 in a single school district in the United States from which we received data.
- (Descriptive statistics versus inferential statistics:) The descriptive statistics of the average height, and related standard deviation, of 12-year-old boys in the sample is a sample statistic that is used to estimate the population parameters of the mean and standard deviation of all 12-year-old boys in the United States.

Since data analysis is a process discipline and statistics is a methodological discipline, both embedded in context, working within these disciplines is not a practice of solving problems, proving theorems, or getting results. It is instead a process of making arguments for certain conclusions based upon the process of observing and analyzing data within a context from which the data is derived. This means that there are no 'right' answers, only strong or weak arguments.

The next few chapters focus on using statistical techniques in the data analysis process. So the distinctions between these two terms will blur as we walk in the overlap between them.

13.1.1 The Centrality of Variability

Variability underlies everything around us, particularly in quantitative situations. As such, variability is the foundation of statistics.

> Individuals vary. Repeated measurements on the same individual vary. In some circumstances, we want to find unusual individuals in an overwhelming mass of data. In others, the focus is on the variation of measurements. In yet others, we want to detect systematic effects against the background noise of individual variation. Statistics provides means for dealing with data that take into account the omnipresence of variability [Cobb and Moore, 1997, p. 801]

Certain situations lend themselves to a deterministic model using mathematical functions. These could include computing the volume of a swimming pool based on certain assumptions about its dimensions. However, estimating the cost of building a swimming pool requires probabilistic models involving statistics. For example, one cannot know exactly how much concrete will be needed for the swimming pool since there is variability in the mixing of the concrete, the effects of temperature and humidity on the application of the concrete, the inability to have perfectly shaped forms used to pour the concrete, and many other factors. So when ordering the concrete for the swimming pool, the contractor needs to take into account the variability of these factors and order an amount of concrete for which he is sufficiently confident he can complete the job. However, he would also not want to order an excess amount of concrete in order to keep down the cost of the project.

The more we learn about the world the more we understand how probabilistic models do a better job of describing our world than deterministic models. For that reason, we predict hurricane paths with a cone of certainty. We can understand how the better sports team lost a game, even though they had an 80% chance of winning.

The goal of statistics is to better understand and quantify the variability in a certain context. We can then interpret and apply this information as we study situations and improve our abilities to make decisions.

13.1.2 Exercises

1. Investigate the origins of the fields of data analysis and statistics and compare them to the history of the field of mathematics.

2. Find 3 unrelated careers that use data analysis as a key aspect of their daily work and describe them.

3. What are some possible contributors to variability in scores on a class exam?

4. With most news reports about the stock market, the rise or fall of the stock index is usually attributed to one or two key news events of the day. This represents a deterministic way of thinking. What would a similar report about the stock market look like that had a more probabilistic way of thinking?

5. Write a short paragraph describing a scenario in which you would use a sample statistic to infer something about a population parameter. Clearly identify the sample, population, statistic, and parameter in your example. Be as specific as possible, and do not use any example discussed in the book or in class.

13.2 Teaching and Learning of Data Analysis and Statistics

Even though the disciplines of statistics and data analysis are very distinct from mathematics in many ways, the teaching and learning of the methods of statistics and the processes of data analysis are often incorporated into the curriculum of mathematics. This incorporation of data analysis and statistics into the field of mathematics education derives from two major factors.

The first is that most disciplines see data analysis as a tool in their discipline, but not one of the central tenants of the discipline. For instance, science uses statistics to account for the reliability of their measurements with probabilistic reasoning taking a more prominent role with concepts such as the make-up of an atom. Similarly, history uses data analysis to understand population changes and make arguments about when and how certain activities like changes in societal structures most likely occurred. However, neither of these disciplines considers data analysis and statistics as a core part of their work.

Organizations such as the National Council of Teachers of Mathematics (NCTM) in documents such as *Curriculum and Evaluation Standards for School Mathematics* [NCTM, 1989] list 'Data Analysis and Probability' as one of the five content strands, thereby actively adopting these disciplines as part of the mathematical sciences.

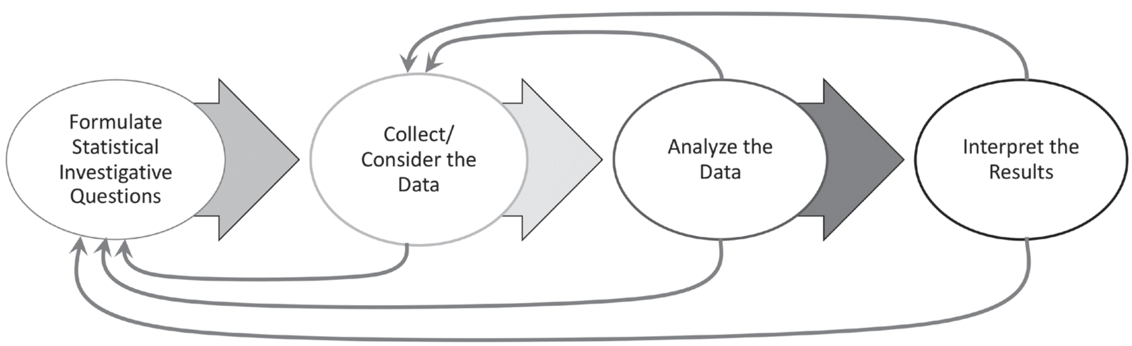


Figure 13.1: Data Analysis Investigative Process [@GAISE2]

In an effort to help teachers of mathematics better understand the discipline of statistics in order that they could teach the subject better, the American Statistical Association published the *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report* [Franklin et al., 2007] and a follow-up report, *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II)* [Bargagliotti et al., 2020]. These documents outline a conceptual framework for statistics education and describes the data analysis process having four major components (formulate questions, collect data, analyze data, and interpret results) that we will discuss in detail below.

While many people think of these four components of the data analysis process as linear, most of the time the process is iterative, jumping between the four components as the work progresses (see Figure 13.1). For instance, a person might produce a graph or chart that causes a question to arise in his or her mind. This question could cause him or her to analyze the data in a different way, modify the interpretation of the results of the current analysis, or seek to collect additional data. Alternatively, a person might notice a pattern when collecting data for some purpose that causes them to make an interpretation, formulate new questions based on the data collected, or change their planned data analysis.

Another important aspect of the data analysis process is the centrality of variability. It is important to remember that "statistical problem solving and decision making depend on understanding, explaining, and quantifying the variability in the data" [Franklin et al., 2007, p. 6]. So the concept of variability needs to be at the forefront of the teacher and students' minds as they work through and learn within the data analysis process.

In order to acknowledge that students have different levels of understanding and skills in the data analysis process, within each of these processes and the underlying focus on variability the GAISE framework considers three different levels (Level A, Level B, and Level C). A teacher needs to be able to recognize where students may be within this framework and support students' progression to higher levels of understanding. Critically, the GAISE framework at no point discusses different types of statistical techniques. This allows this same framework to be used in the teaching and learning of data analysis for sixth graders learning how to use box-and-whisker plots or high school seniors learning how to use a χ^2 test for categorical data.

In order to understand this data analysis investigative process and the centrality of variability within the process we will give a more detailed explanation of each of the components, along with examples highlighting the different levels within that process component. For a more detailed description, see the full GAISE Reports [Franklin et al., 2007, Bargagliotti et al., 2020].

Related Content Standards

- (6.SPA.1) Recognize a statistical question as one that anticipates variability in the data related to the question and accounts for it in the answers.

Formulate Questions. A key component of formulating questions in the data analysis process is having questions that anticipate variability.

- **Level A.** A student or class is at the first level of the framework if the teacher is the one posing the questions of interest or is having to reword students' questions from those with a deterministic or definite answer, such as "How many desks are in this classroom?", to become questions that allow for, and expect, variability, such as "How many students are in most classrooms in our school district?".
- **Level B.** The second level corresponds with the ability to formulate questions that account for sampling variability within a group or between groups. The questions also begin to recognize the distinction between a population and a sample. Some sample questions would be, "Are students in our class generally taller than students in the other classes in the school?" or "Are 14 year old boys usually shorter than 14 year old girls?".
- **Level C.** Students at Level C pose questions that also account for chance variability and think of the world in a more probabilistic sense. These questions often include questions about the strength of relationships and/or the reliability of the measurements creating the data.

Collect Data. The design of the data collection component of the investigative process should acknowledge that variability exists, take steps to reduce the amount of variability due to factors that are not the main purpose of the investigation (such as measurement errors), and use techniques such as random sampling in order to reduce the differences between the sample and the population. A key component of this data collection design process involves the affect of the sample size upon the reliability of the results.

- **Level A.** These individuals do not account for sample size (it might be mentioned, but no connections are made with the other components of the process), ignores possible measurement errors, and does not use random selection.
- **Level B.** This level involves accounting for some of the aspects of controlling variability in the data collection process mentioned above, but does not account for all of them.
- **Level C.** Individuals at this level explicitly account for sample size, measurement error, and uses random sampling techniques when using samples to predict properties of a larger population.

Analyze Data. The purpose of the analysis component in the investigative process is to give an accounting for the variability in the situation.

- **Level A.** These students are able to create graphical displays, compute parameters, or perform statistical tests on data, but are unable to explain why a certain analysis technique is used. Such a student may find the mean and standard deviation of a set of data, but does not explain why those parameters address the situation better than a median and interquartile range.
- **Level B.** A Level B understanding corresponds with the quantification of the variability, but uses a single technique or does not consider the underlying distribution in the justification of the techniques used.
- **Level C.** A Level C student focuses on using multiple techniques to build an argument based in the context of the data. The analyses also recognize the underlying probability distributions involved in the situation as part of the justification of the statistical techniques used.

Interpret Results. Interpretation must be intertwined with an understanding of the context of the data and attempt to explain the sources of the variability in the situation. Quality interpretations must take into account the sample size; effects of random selection; measures of strength of association and models; distinguish between causation and correlation; and the difference between different types of studies.

- **Level A.** A Level A interpretation gives some results and states a conclusion without discussing the other aspects of the situation. These students focus on the data and struggle to generalize to the larger context.
- **Level B.** These interpretations move beyond basic summaries, but do not account for all of the variability and nuance of the context.
- **Level C.** A Level C interpretation discusses how the different techniques are used and the results from the analysis work together to justify the conclusions. They also keep in central the context of the situation and the variability involved. }

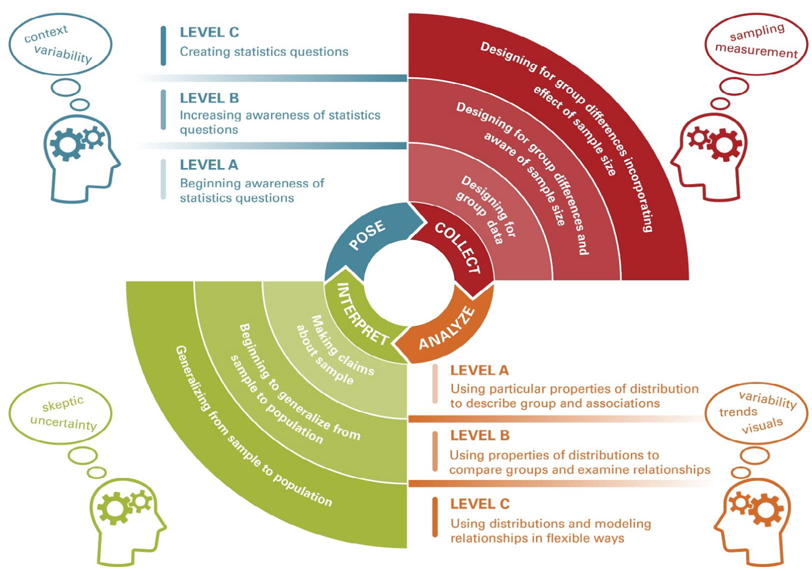


Figure 13.2: Students' Approaches to Statistical Investigations (SASI) Framework @SASI

13.2.1 Statistical Habits of Mind

As students engage in the data analysis investigative process, one of the primary goals for students is to develop a new way of thinking about the world and the use of data in our world. Chance [2002] listed six mental habits for students to develop as they learn to think statistically. “These mental habits include:

1. consideration of how to best obtain meaningful and relevant data to answer the question at hand
2. constant reflection on the variables involved and curiosity for other ways of examining and thinking about the data and problem at hand
3. seeing the complete process with constant revision of each component
4. omnipresent skepticism about the data obtained
5. constant relation of the data to the context of the problem and interpretation of the conclusions in non-statistical terms
6. thinking beyond the textbook"

The chapters on data analysis in this text are designed to help develop these habits of mind through examples and engagement in various projects involving the investigative process. Remember that in the teaching and learning of probability, data analysis, and statistics the main goal is not to learn how to perform statistical calculations, but to develop these habits of mind and improve one's ability to produce and understand arguments based upon the analysis of data.

13.2.2 Exercises

1. Write a short essay on why data analysis should be taught in K-12 schools.

2. Determine where each of the following questions fall in the framework. Then modify the question to be at a Level C.

   a. How many people are in this classroom?
   b. What type of food do students in our school prefer?
   c. What types of memes spread fastest?

3. Determine where each of the following scenarios for the collection of data would fall in the framework. Then modify the scenario to be at a Level C.

   a. In an effort to determine if plants grow better in red light or blue light a student has one plant that grows under a red light bulb for 4 weeks and one plant that grows under a blue light bulb for 4 weeks. The student then measures the difference in the heights of the plants.
   b. The students in a class measure their height and arm span in order to determine if there is a relationship between the two measurements for people.
   c. Using the nutrition information for McDonald's, students look for how many calories are in a gram of fat.

4. Determine where each of the following scenarios for the analysis of data would fall in the framework. Then modify the scenario to be at a Level C.

   a. In order to determine how much it will cost to go to college, students find the mean tuition rates of all of the colleges in their state.
   b. In order to predict who will win a presidential election, a poll asks a random sample of likely voters across the country who they will vote for. This data is analyzed, giving a percentage of votes for each person with a margin of error.

5. Determine where each of the following scenarios for the interpretation component of the investigative process would fall in the framework. Then modify the scenario to be at a Level C.

   a. From 2017-2019 the information about fatal police shootings[1] indicated that 1,226 white people, 667 black people, and 485 Hispanic people were fatally shot by police. A person concludes that a white person is more likely to be fatally shot by police than a black person.

[1] https://www.statista.com/statistics/585152/people-shot-to-death-by-us-police-by-race/ retrieved August 1, 2020

6. Find a news article or research article of interest to you that involves a data analysis process and analyze the article within the context of the GAISE Framework to describe the probable level within each of the four processes represented by the article.

Chapter 14

Exploring Data

14.1 Types of Data

A core component in the data analysis process is data that can be in many different forms. When we collect or analyze data it is usually made up of **cases** that are the objects in the collection that are the intended **unit of analysis**. Sometimes these objects are the students in a classroom, houses in a neighborhood, or individuals who filled out a survey. The cases are often labeled with some type of name or number to distinguish between different cases. Sometimes the data has been changed to instead give the number of cases with a certain property, rather than listing each case individually. For example, we know that in our classroom that the favorite color of 8 students is red, 7 students like green, 2 students like blue, and 4 students prefer purple.

For each of the cases in our data set, there corresponds one or more attributes, called **variables**. These variables can be verbal or numerical descriptions of some property that varies among the cases studied. Using a **code book** is extremely useful to keep track of the cases in a study and the variables included.

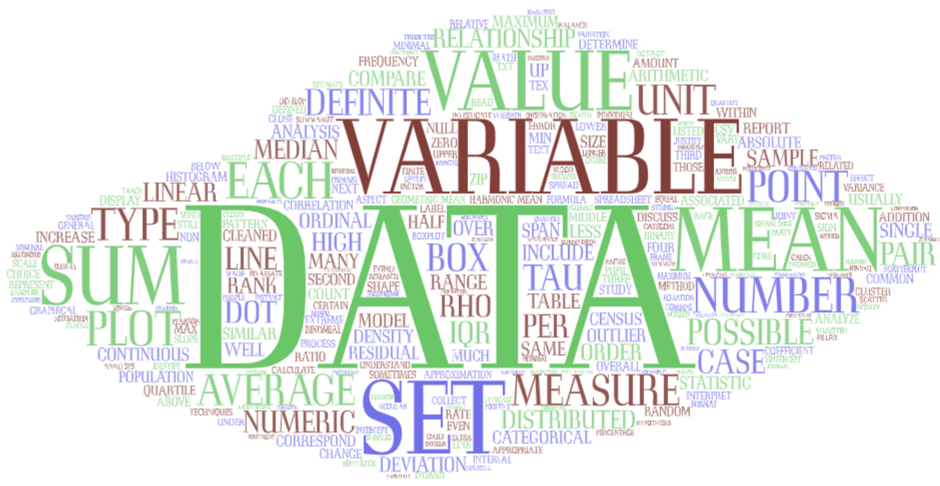
Example 14.1. At various points throughout this course we will be using the `mosaic` package in R, along with the mosaic data sets in the `mosaicData` package. Here is the code book for one of the data sets that includes SAT scores. While the type of variable is not listed, it can easily be inferred from the variable description.

State by State SAT data

- *Description:* SAT data assembled for a statistics education journal article on the link between SAT scores and measures of educational expenditures

- *Format:* A data frame with 50 observations on the following variables.

  - `state` a factor with names of each state
  - `expend` expenditure per pupil in average daily attendance in public elementary and secondary schools, 1994-95 (in thousands of US dollars)
  - `ratio` average pupil/teacher ratio in public elementary and secondary schools, Fall 1994
  - `salary` estimated average annual salary of teachers in public elementary and secondary schools, 1994-95 (in thousands of US dollars)
  - `frac` percentage of all eligible students taking the SAT, 1994-95
  - `verbal` average verbal SAT score, 1994-95
  - `math` average math SAT score, 1994-95
  - `sat` average total SAT score, 1994-95


DATA
VARIABLE
VALUE
MEAN
SUM
SET
NUMBER
EACH
AVERAGE
MEASURE
POINT
PLOT
TYPE
UNIT
DEFINITE
RELATIONSHIP
POSSIBLE
BOX
TAU
RHO
MEDIAN
LINEAR
CASE
DOT
PAIR
SAMPLE
DISTRIBUTED
NUMERIC
LINE
PER
IQR
RANK
HIGH
MANY
ORDINAL
COMPARE
ARITHMETIC
MAXIMUM
RELATIVE
FREQUENCY
ANALYSIS
HISTOGRAM
CORRELATION
CONTINUOUS
POPULATION
QUARTILE
CATEGORICAL
DEVIATION
CORRESPOND
CHANGE
RESIDUAL
DENSITY
MODEL
RANGE
TABLE
INCLUDE
SPAN
CENSUS
OUTLIER
ORDER
STATISTIC
ABSOLUTE
REPORT
SINGLE
ADDITION
COMMON
ANALYZE
INCREASE
DISPLAY
SIMILAR
WELL
COUNT
SECOND
CLEANED
SHAPE
SAME
MUCH
STUDY
USUALLY
VARIANCE
ASSOCIATED
GEOMETRIC MEAN
HARMONIC MEAN
FORMULA
SPREADSHEET
APPROXIMATION
OVERALL
RANDOM
INTERPRET
SOMETIMES
UNDERSTAND
CALCULATE
ABOVE
TECHNIQUES
INTERVAL
APPROPRIATE
RATIO
PROCESS
BINOMIAL
EXTREME
CERTAIN
NULL
ZERO
UPPER
LOWER
SIZE
MIN
TEXT
UP
TEX
MATH
AMOUNT
DETERMINE
MINIMAL
WITHIN
LISTED
THIRD
THOSE
RELATED
NEXT
ZIP
HALF
OVER
LABEL
BOXPLOT
MIDDLE
LESS
DISCUSS
FOUR
METHOD
BINARY
THREE
SIGN
CLUSTER
SCATTERPLOT
SCALE
CHOICE
REPRESENT
GENERAL
GRAPHICAL
MAX
SLOPE
PATTERN
NON
STILL
RATE
EVEN
COLLECT
BELOW
ASPECT

The code book is usually written in some type of text document (often a `.txt` file) that can be read by anyone that may later want to use the data set. The data for an investigation is usually cleaned and stored in a spreadsheet (often in a `.csv` format to make it easier to transfer between software for analysis). By having the code book stored with the spreadsheet, the spreadsheet can have all of the extra information removed to make it easier to use for analysis.

In the code book we often describe the cases, including the number of cases, and the techniques used to collect the data. The variables for the cases are then each listed with their variable name (usually a single word or a string of words connected by underscores, i.e. `arm_span`), a description of the property described by the variable, possible values for the variable and their meanings (for example when describing educational level of parents we sometimes use 0= did not graduate high school, 1= high school diploma, 2=some college, 3= college degree), and the units of measurement (if applicable). You would also want to include any additional information about the variable that you may need to remember as you work through the data analysis process. Remember that the context and information about the cases and variables need to be in the forefront of your mind throughout the entire investigative process.

Related Content Standards

- (6.SPB.5) Summarize numerical data sets in relation to their context, such as by:

  a. Reporting the number of observations.
  b. Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

A key aspect of the variables that should be included in the code book is the type of data involved, as the type of data impacts the types of statistical techniques that can be applied.

14.1.1 Categorical

A variable that can be put into a finite number of categories such as color, blood type, political party affiliation, zip code, or gender. These categories descriptions are often each replaced with a single word or number in the corresponding data spreadsheet to allow for the statistical software to run analyses. The correspondence between these abbreviations and the original description is included in the code book for the data set.

Since categorical data does not have an intrinsic ordering to it, it is called **nominal**.

14.1.2 Ordinal

A variable that has some type of order within the possible values of the variable, but the differences between successive values is imprecise. A common example is the ranking of college football teams. While there is an order in the ranking of the teams, the difference between the second and third ranked teams is not likely to be the same as the difference between the third and fourth ranked teams. Another such example is the order in which runners finish a race. There may only be 1 second between the first and second place finishers, while the difference between the seventh and eighth place finishers may be 20 seconds.

When analyzing ordinal data, the order is usually entered as an integer value, so it is important to label the variable as ordinal, rather than a count or continuous, as the types of analyses run on ordinal data is very different than other types of variables. Such analyses usually use non-parametric statistical techniques since the data does not fit within a normal distribution.

14.1.3 Binary

A variable that has two possible values associated to it. These are often 'yes/no', 'true/false', or 'correct/incorrect' type values. When stored in the corresponding spreadsheet, these are usually replaced with '0/1' options with the correspondence between the number and the actual value of the variable described in the code book.

Sometimes binary variables have an order (like correct or incorrect on a test question), in which case they are a type of ordinal variable. Other times there is no order to the values of the variable and so it can be thought of as a categorical variable.

14.1.4 Binomial

A variable based on the number of successes out of N possible. A common example of a binomial variable is the number of heads achieved when flipping a coin 10 times. A more complex example is the number of patients with a disease from a sample of 15 patients randomly chosen from different hospitals. In this situation, the hospital is the case and the number of patients (out of 15) is the binomial variable.

14.1.5 Count

A count is very similar to a binomial variable, but it is not limited to a certain number possible. It could be something like the number of people standing in line at different check-out stations in a supermarket or the number of kids in each classroom of a school.

14.1.6 Continuous

When the value of a variable can range over a large number of possible values where the differences in values have meaning, then the variable is assumed to be continuous, or real-valued. These could include temperatures, test scores, heights, or speeds.

14.1.7 Exercise

1. Determine a variable name, type(s) of variable, units of measurement, and possible values for each of the following descriptions. (not all of these are well defined and so you will need to make a case for some of your choices)

   a. age
   b. amount of time spent interacting with a screen each day
   c. number of views of a YouTube video
   d. height of students
   e. calories in a hamburger
   f. if a person is voting for a certain candidate
   g. number of wins for a football team during a season

2. Consider the following variables from the Census at School United States[1] data:

   - Region: Identifies the state the participant lives in. (50 possible values, 1 for each state)
   - Planned education level: Indicates the highest degree a student intends to earn. (6 possible values: less than high school, high school, some college, undergraduate degree, graduate degree, other.)

[1] https://ww2.amstat.org/censusatschool/

- Reaction time: The amount of time, in seconds, it takes to click their mouse after an image appears on a screen. (Range is theoretically 0—not inclusive– to infinity.)
- Memory game score: Score on a memory assessment, with the score corresponding to the number of moves it takes to solve a memory puzzle. (Minimum score = 20, no theoretical maximum.)
- Favorite season: Name of student's favorite season. (4 possible values.)
- School work pressure: The amount of pressure the students identify as experiencing in response to the question "How much pressure do you feel because of the schoolwork you have to do?" (4 possible values: none, very little, some, a lot)

Determine the types of variables for each of these.

14.2 Exploring Univariate Data Graphically

The data analysis investigative process often begins by a person exploring an existing data set looking for possible phenomena or relationships that lead to further questions to be tested using a more systematic statistical analysis. Both the exploration of relationships and the confirmation of hypothesized relationships are valuable, but they serve different purposes.

The primary goals of **exploratory data analysis** (EDA) revolve around getting to know a set of data and allowing the data to guide further discoveries. Behrens and Yu [2003] describe it as similar to a "detective looking for clues to develop hunches and perhaps seek a grand jury" (p. 42). EDA uses many tools, particularly graphical representations, to reveal possible relationships, understand error and noise, transform variables, and better understand the role of outliers. Throughout this section we will describe these different tools and how they are used in spreadsheets and R.

> In the early stages of research, EDA is valuable to help find the unexpected, refine hypotheses, and appropriately plan future work. In the later confirmatory stages, EDA is valuable to ensure that the researcher is not fooled by misleading aspects of the confirmatory models or unexpected and anomalous data patterns. [Behrens and Yu, 2003, p. 60]

As an example of the process of exploratory data analysis and different ways of representing a set of data, we will look at the time between geyser eruptions of Old Faithful for July 2020[2]. Using the information downloaded about the time of eruptions, we create a column in a spreadsheet called `inter_eruption_time` that is calculated based on the time of the eruption and the time of the prior eruption in minutes.

14.2.1 Histograms

We will begin our exploration of the Old Faithful data with a histogram with a binwidth of 15 minutes to show the number of minutes between eruptions.

Using Excel or Google Sheets we can highlight the column and insert a histogram.

Alternatively we can use R and the `readxl` and `ggplot2` packages to create a histogram.

```
# Load required packages
library(readxl)
library(ggplot2)
```

After loading the appropriate packages we can read the Excel file into the data frame of `Old_Faithful_2020_07_01_to_2020_07_31`. This can then be used to create the histogram with the binwidth of 15 minutes.

[2]Downloaded from https://geysertimes.org/retrieve.php

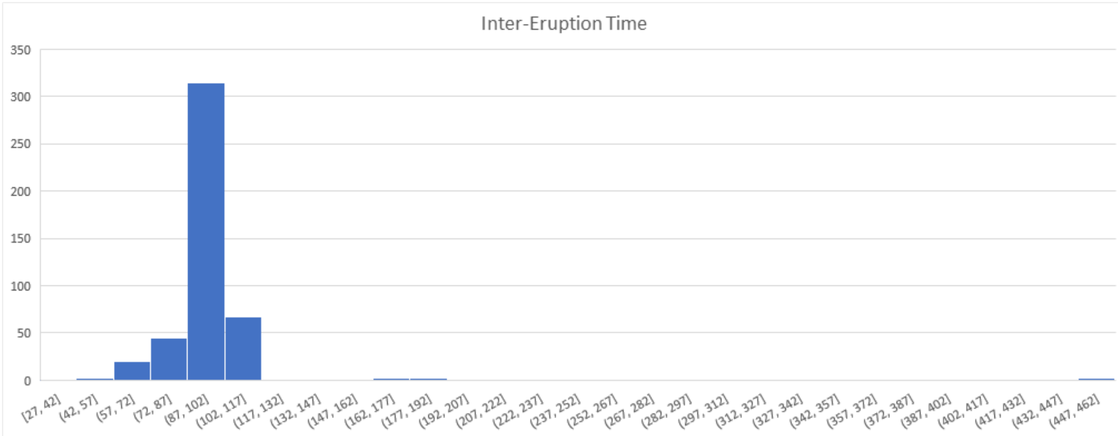
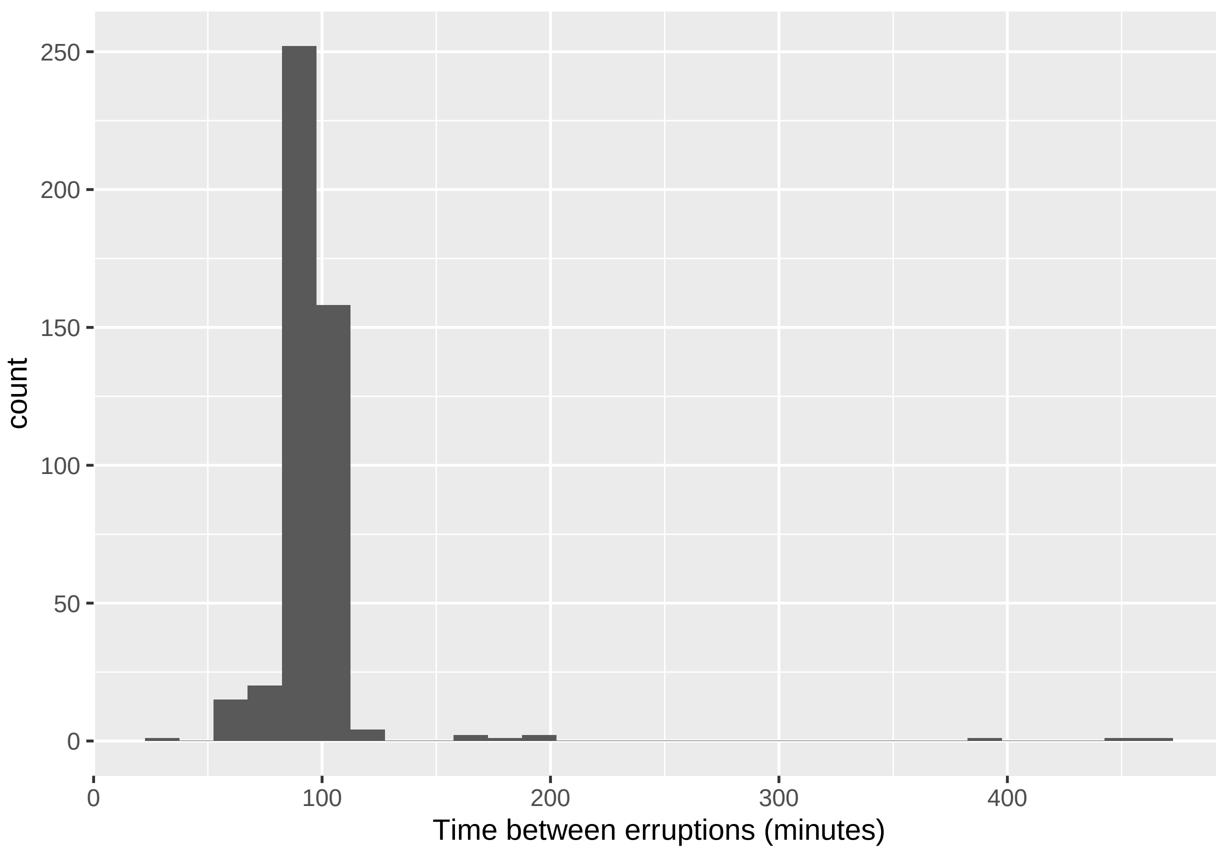


Figure 14.1: Histogram with 15 minute binwidth using Excel

```
ggplot(Old_Faithful_2020_07_01_to_2020_07_31, aes(x=inter_erruption_time)) + geom_histogram(binwidth = 
```



We notice that the vast majority of the time, the number of minutes between eruptions is less than 200 minutes, with a few over 400 minutes. Going back to the context of the data and looking at these eruptions that had more than 120 minutes, they all occurred during nighttime hours. So we will make the assumption that the data source is missing eruptions and so we will remove these data points from our frame. We then create a new histogram with a binwidth of 5 minutes on this cleaned data set in both Excel and R.

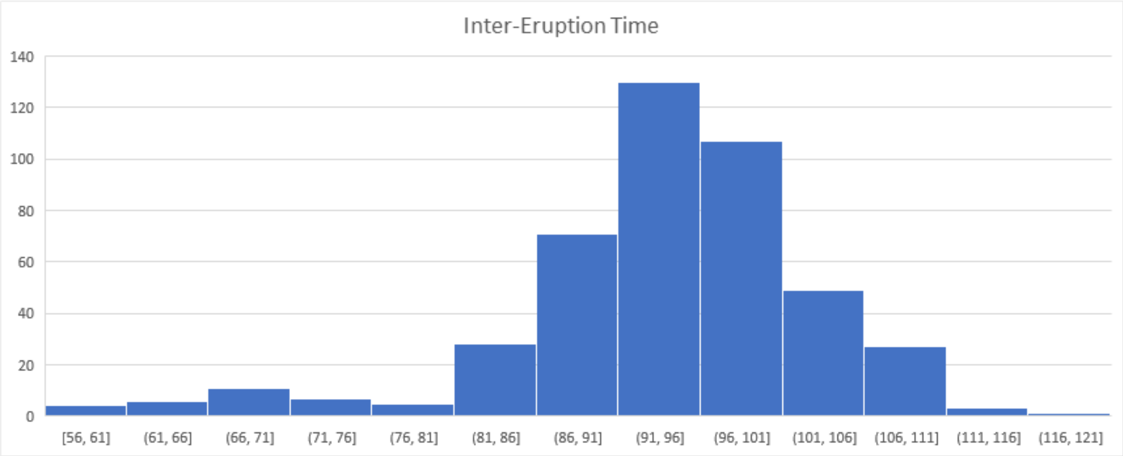


Figure 14.2: Histogram with 5 minute binwidth using Excel

```
ggplot(Old_Faithful_2020_07_01_to_2020_07_31_cleaned, aes(x=inter_erruption_time)) + geom_histogram(bin
```

Related Content Standards

- (6.SPB.4) Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

14.2.2 Dot Plots

From this histogram we see that the data seems to have two 'centers'. One at around 95 minutes and the other around 65 minutes. To explore this phenomenon further, we create a dot plot that marks each occurance as a single dot and see that there appear to be two clusters of times.

```
ggplot(Old_Faithful_2020_07_01_to_2020_07_31_cleaned, aes(x=inter_erruption_time)) + geom_dotplot(binwi
```

While spreadsheet applications do not create such dot plots easily, you can use GeoGebra or Desmos to create dot plots.

14.2.3 Density Plots

Since the dot plot represents the information from a single month and the times between eruptions is a continuous variable, we can generalize the dot plot to a continuous density plot.

```
ggplot(Old_Faithful_2020_07_01_to_2020_07_31_cleaned, aes(x=inter_erruption_time)) + geom_density(kerne
```

14.2.4 Box Plots

When we display this same data on a box plot (also called a box-and-whisker plot), we see that many of the data points lie more than 1.5 times the inner-quartile range (IQR) below the first quartile (Q1) or above the third quartile (Q3), as denoted by the dots past the whiskers. These data points are often called outliers.

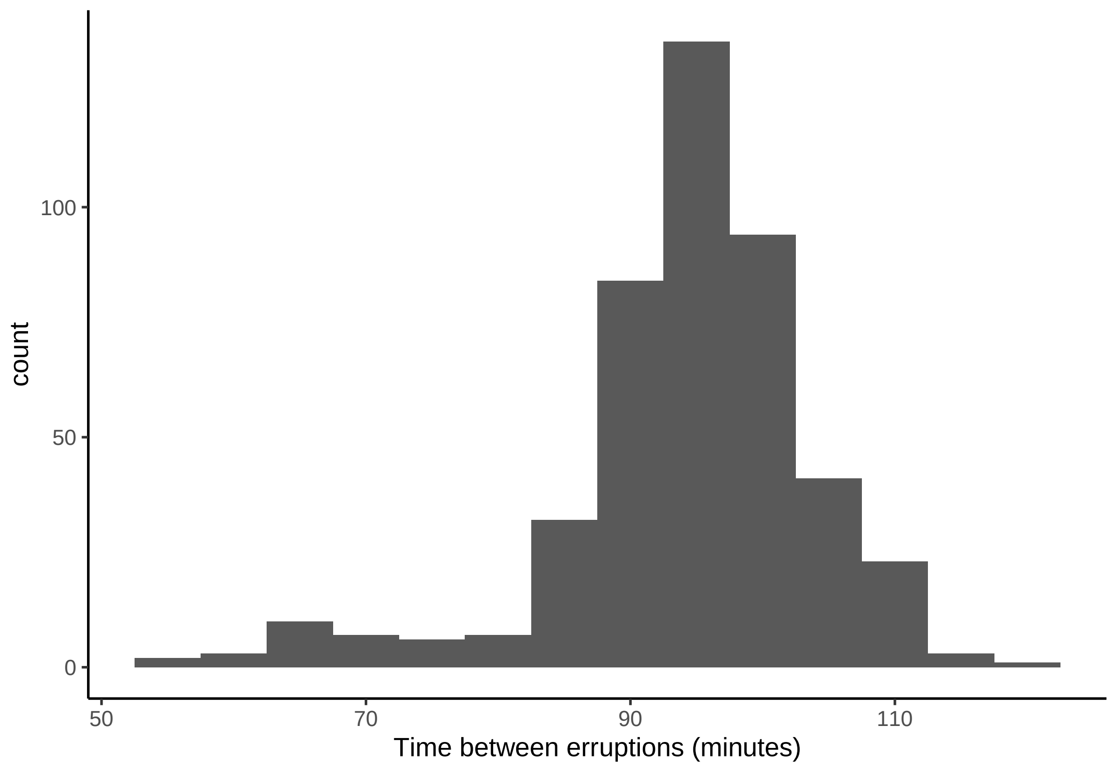


Figure 14.3: Histogram with 5 minute binwidth using R and ggplot2

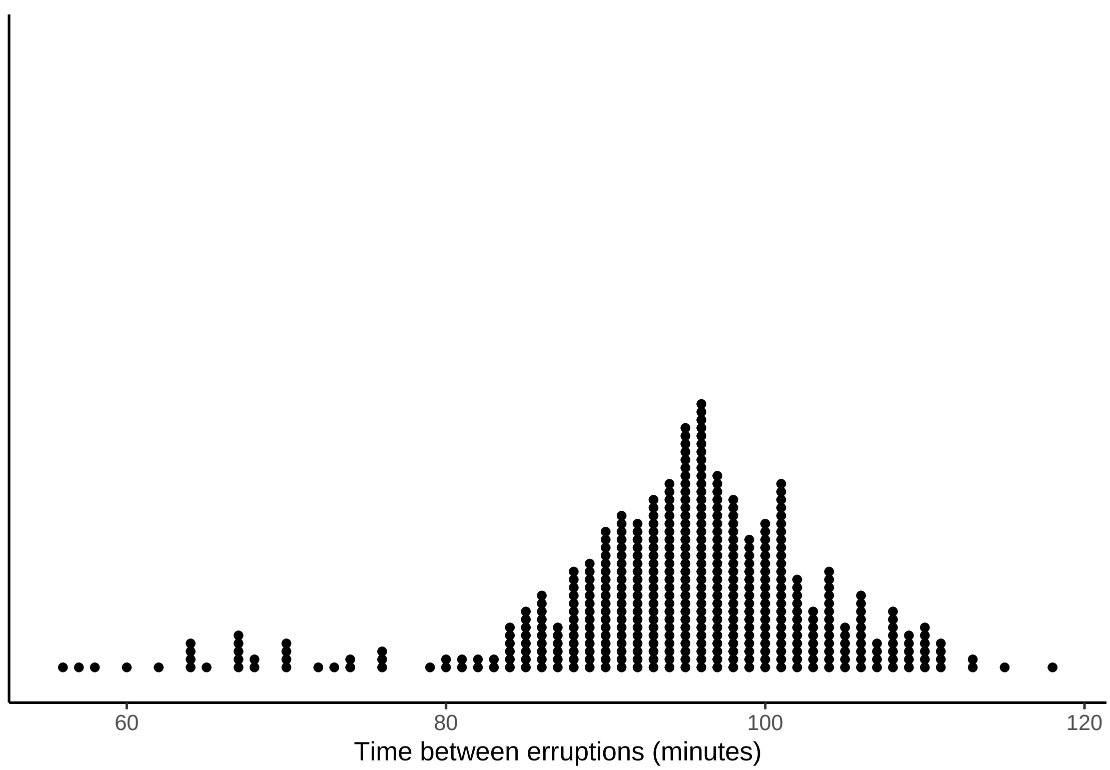


Figure 14.4: Dot Plot of Eruption Times using R and ggplot2

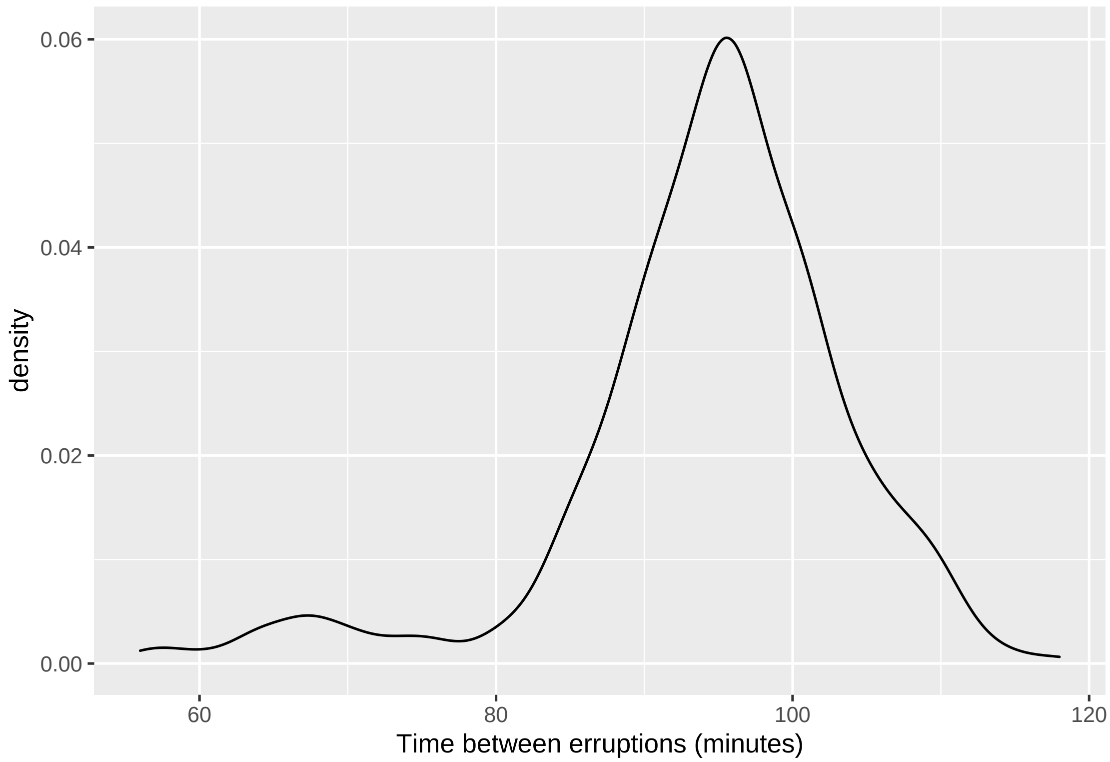


Figure 14.5: Density Plot of Eruption Times

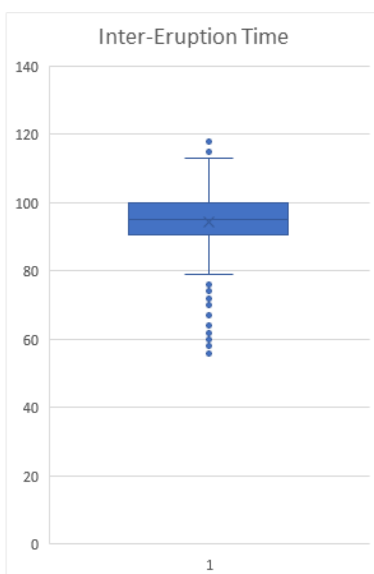


Figure 14.6: Box Plot of Time Between Eruptions Using Excel

```
ggplot(Old_Faithful_2020_07_01_to_2020_07_31_cleaned, aes(x=inter_erruption_time)) + geom_boxplot() + la
```

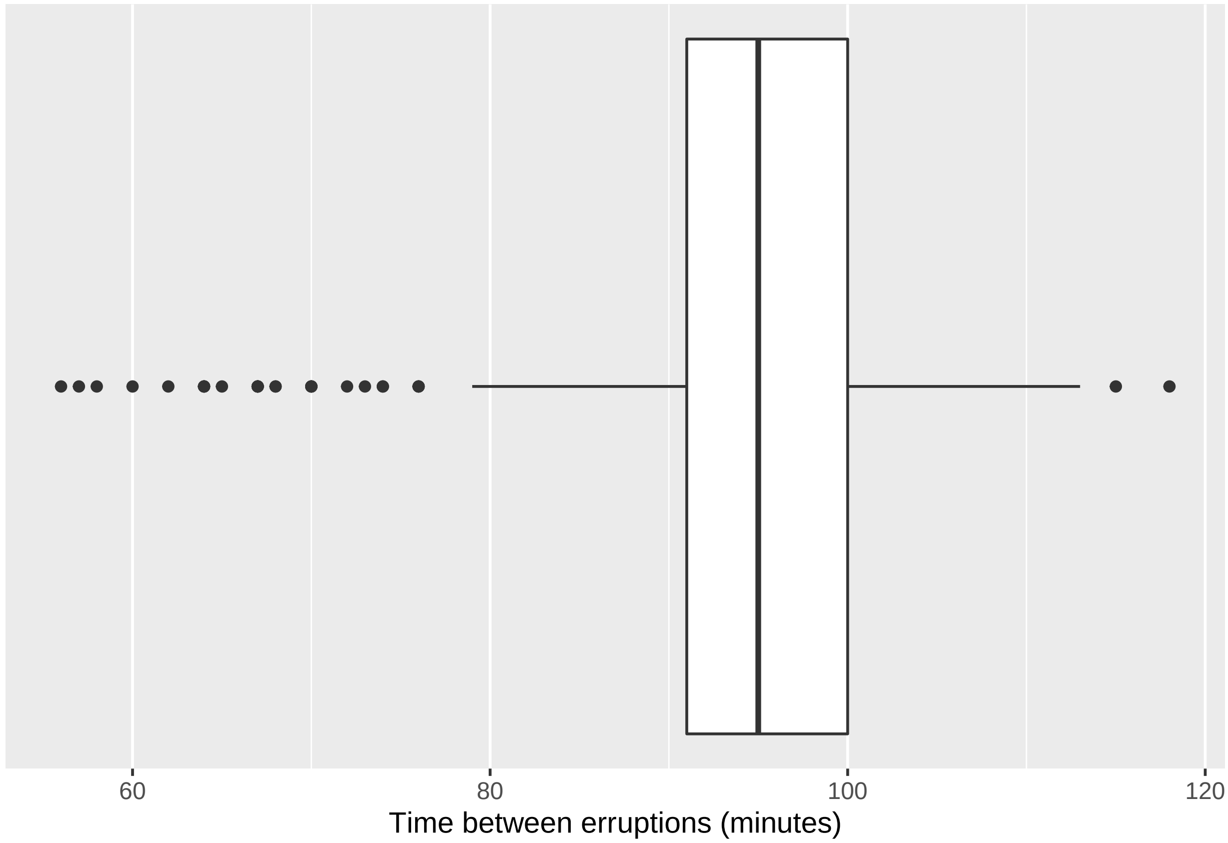


Figure 14.7: Box Plot of Eruption Times Using R and ggplots2

The large number of outliers below 78 minutes corresponds to the second cluster of times centered around 65. From the evidence gathered to this point, we can justify making the assumption that there are really two different lengths of time between eruptions, "Short" and "Long". By labeling the wait time as short or long, we can look at the two different distributions using a pair of box plots.

```
ggplot(Old_Faithful_2020_07_01_to_2020_07_31_cleaned, aes(inter_erruption_time, wait_cat)) + geom_boxpl
```

14.2.5 Exercises

1. Use the Census at School Random Sampler https://ww2.amstat.org/censusatschool/ to create a spreadsheet with 1000 random students. Use this data set to explore the variables of `Gender`, `Age_years`, `Height_cm`, and `Armspan_cm` using the techniques discussed in this section. Create a report to discuss your findings.

2. Use the tools from this section to explore the sugar content of regular soft drinks, juices, milk, and sports drinks.

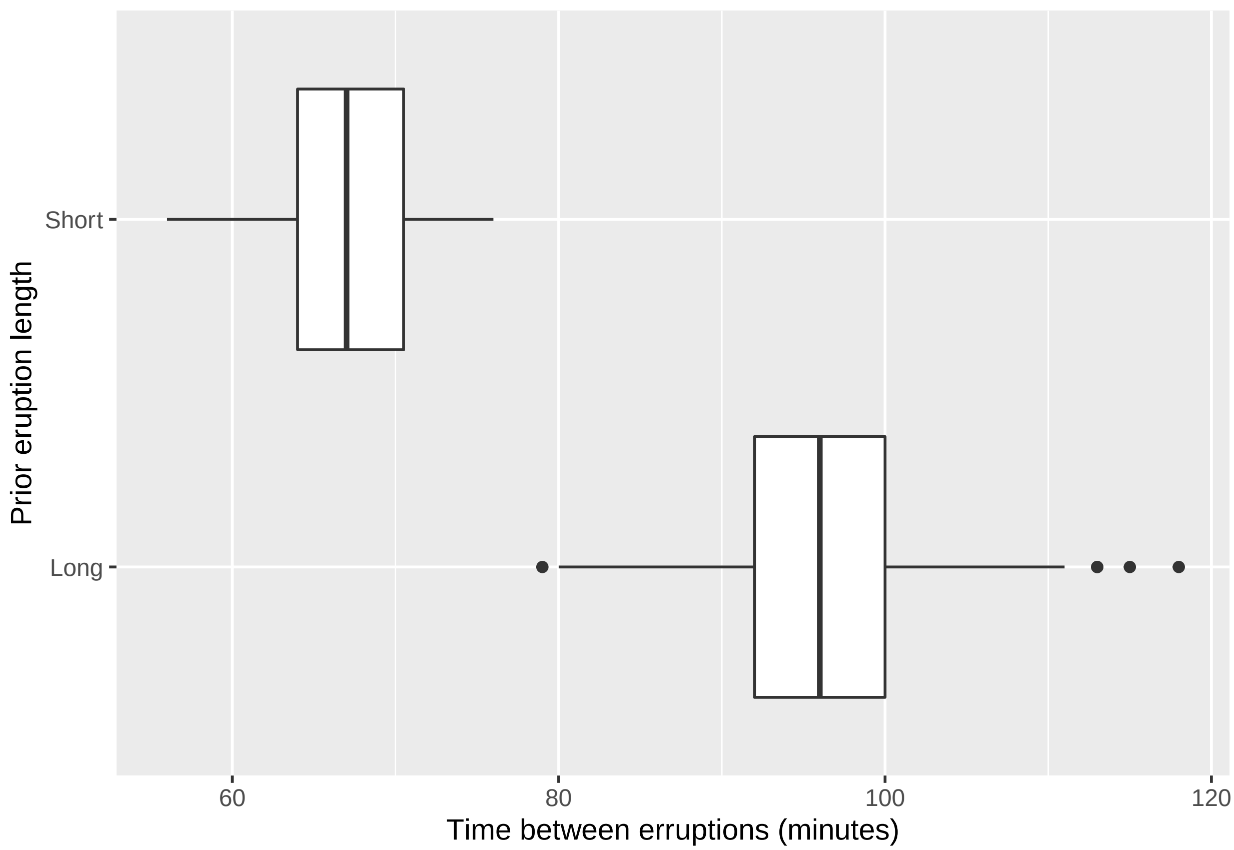


Figure 14.8: Box Plot of Eruption Times Based on Prior Eruption Using R and ggplots2

14.3 Exploring Bivariate Data Graphically

Related Content Standards

- (7.SPB.3) Informally assess the degree of visual overlap of two numerical data distributions with similar variabilities, measuring the difference between the centers by expressing it as a multiple of a measure of variability.

In addition to using each of the graphical displays shown above to understand a single variable, we can use these displays to compare the distributions for different subsets of the population. For example, we will compare the time between eruptions for Old Faithful with two other geysers in Yellowstone National Park, Daisy and Riverside.

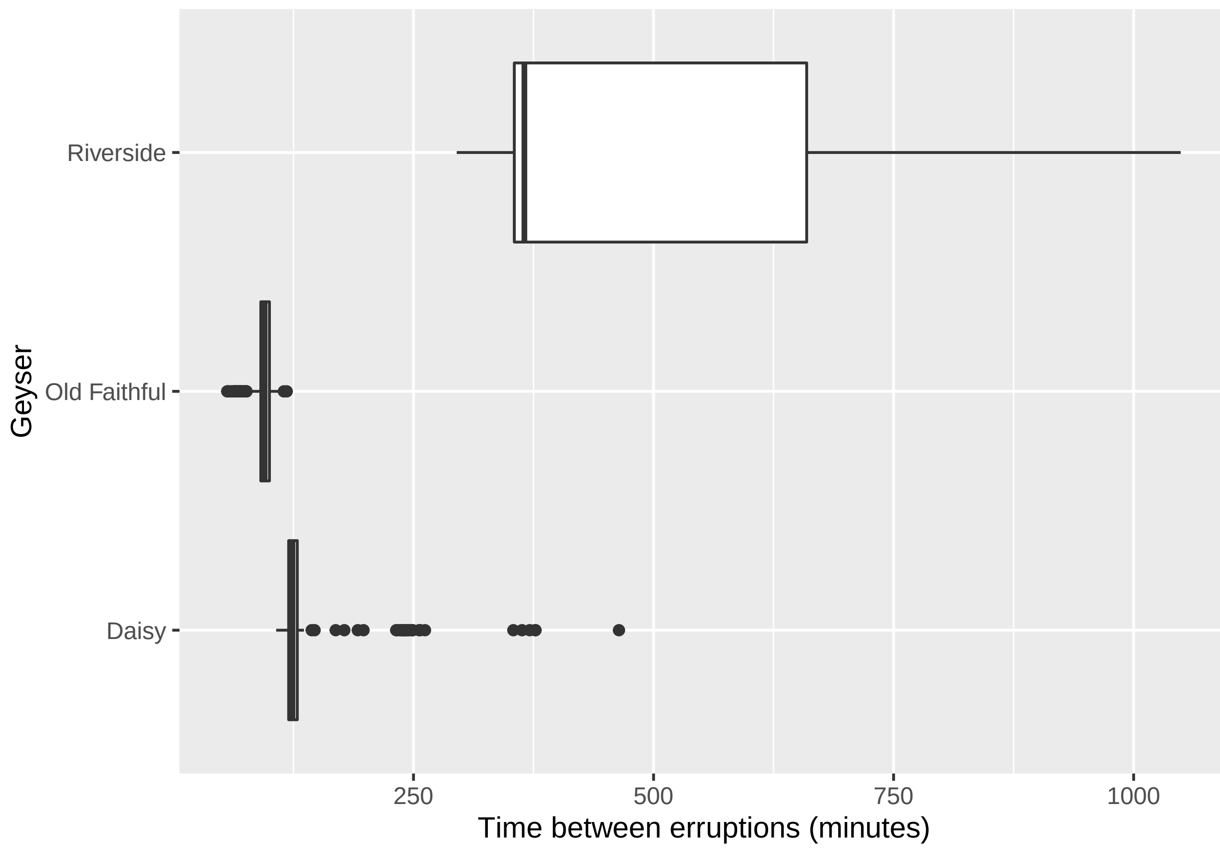


Figure 14.9: Box Plot Comparing Erruption Times of Geysers

We can see from a comparison of their box plots that Old Faithful is definitely the most regular of the three geysers. But Daisy is also very regular, with a little longer of time between eruptions. Some of the longer lengths in eruptions for Daisy may be due to a missed eruption in the recordings. Particularly since it is not as popular of a geyser, not all of its eruptions may have been recorded. This hypothesis is further supported by clusters of outliers around two and three times the length of the eruptions for Daisy and Riverside is highly skewed right. This exploratory data analysis would then lead to a further study of the geysers with more reliable methodology to determine the time between geyser eruptions.

These graphical representations are helpful to discover relationships between an ordinal or continuous variable and a categorical variable. In the displays above, we can think of the name of the geyser as a categorical variable and the time between eruptions as a continuous variable.

14.3.1 Scatterplots

It is sometimes useful to see how two continuous or ordinal variables interact with each other. To explore this interaction, a scatterplot of the data is often very helpful. For some of the eruptions of Old Faithful in July 2020 we have the length of the eruption recorded. With this additional information we create a scatterplot of the length of the previous eruption and the time between eruptions.

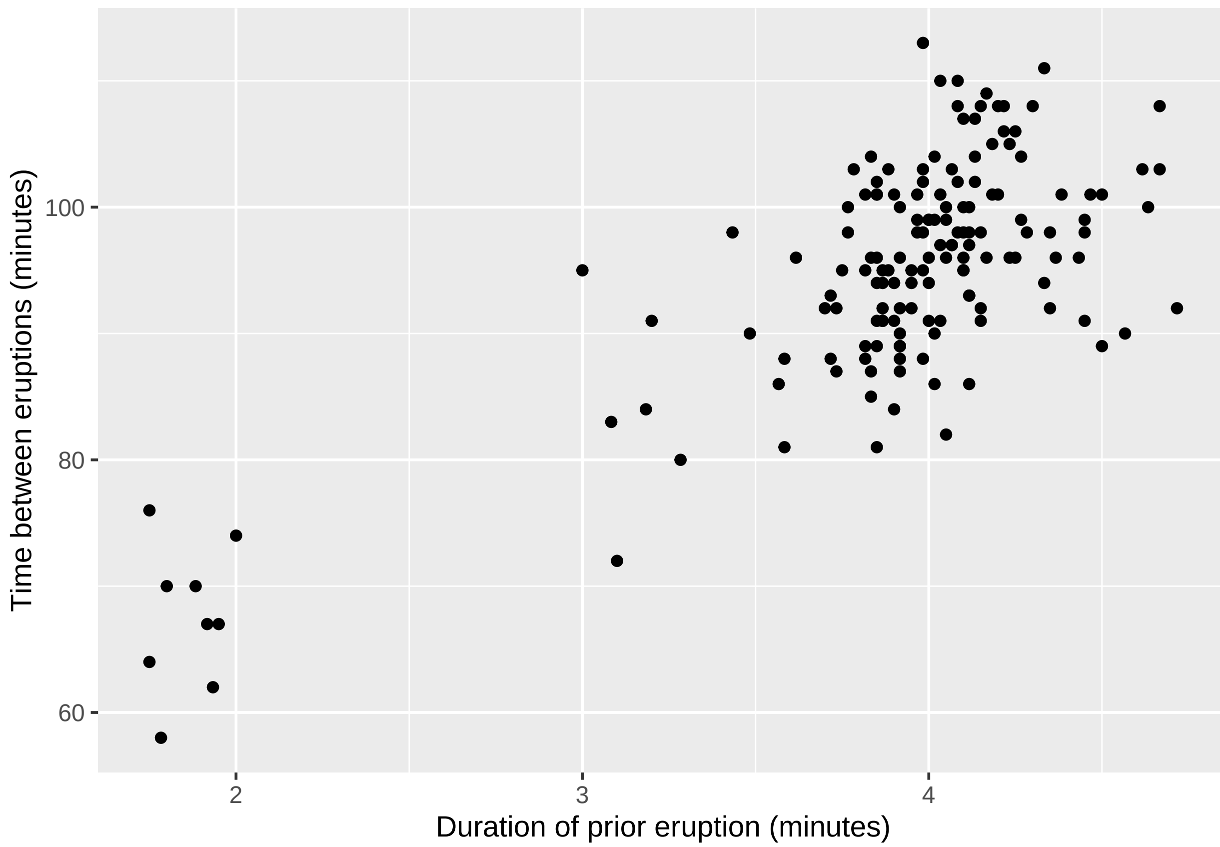


Figure 14.10: Scatterplot Comparing Erruption the Duration of Prior Eruptions and the Times Between Erruptions (R)

From this scatterplot we can see two main clusters of eruptions, those with around 2 minute long eruptions that then have around 70 minutes until the next eruption and those with four minute eruptions with the next eruption around 90 minutes later. This could lead to a hypothesis that the length of an eruption influences the time until the next eruption. Remember that we cannot say anything definite here. Instead, we can create hypotheses and more detailed research plans to build upon these exploratory analyses to develop a more rigorous argument.

Related Content Standards

- (8.SPA.1) Construct and interpret scatter plots for bivariate measurement data to investigate patterns of association between two quantities. Describe patterns such as clustering, outliers, positive or negative association, linear association, and nonlinear association.

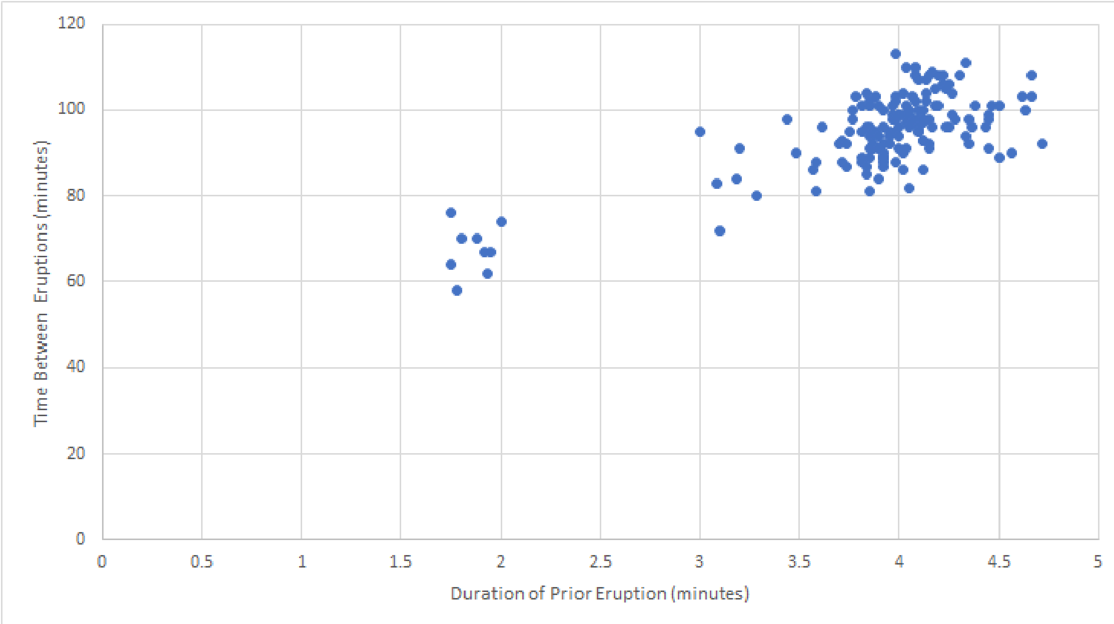


Figure 14.11: Scatterplot Comparing Erruption the Duration of Prior Eruptions and the Times Between Erruptions (Excel)

14.3.2 Exercises

1. Use the Census at School Random Sampler https://ww2.amstat.org/censusatschool/ to create a spreadsheet with 1000 random students. Use this data set to explore possible relationships between pairs of the variables of `Gender`, `Age_years`, `Height_cm`, and `Armspan_cm` using the techniques discussed in this section. Create a report to discuss your findings.
2. Use the tools from this section to explore possible relationships between pairs of variables of nutrition information of regular soft drinks, juices, milk, and sports drinks.

14.4 Measures of Center

Many of the graphical representations from the previous sections help to illuminate properties of the distribution of data along a single numerical variable. This section will provide ways to describe properties of such distributions numerically. These numerical descriptions provide additional ways to describe a distribution's center, spread, and overall shape.

Related Content Standards

- (6.SPA.2) Understand that a set of data collected to answer a statistical question has a distribution which can be described by its center, spread, and overall shape.
- (6.SPA.3) Recognize that a measure of center for a numerical data set summarizes all of its values with a single number, while a measure of variation describes how its values vary with a single number.

One of the first attributes of a set of numerical data that most people want to know is a single number that describes the 'average' or 'center' of the data. Consider the following two questions:

1. What is the average height of students in 6th grade?

2. What is the height of the average student in 6th grade?

How are the questions similar and how are they different?

In addition to differences between wording, there are different ways to calculate the 'average'. Consider the following three examples:

- If a class of 25 students has 6 packages of cookies, with each package having 20 cookies, what is the average number of cookies that each person receives if the cookies are distributed evenly?
- A motorboat makes a 24 mile upstream trip on a river against the current in 3 hours. The returning trip using the same amount of propeller rpm takes 2 hours. What is the motorboat's average speed?
- Your school was given a painting worth \$5,000 4 years ago. The painting increased in value by 50% the first year, 20% the second year, and decreased by 10% the third year. It then increased by 5% this year. What is the average annual percentage increase over the 4 years?

With each of these three examples we are calculating a different type of average, or mean. Each of these means are equally based in mathematics and have usefulness in analyzing data.

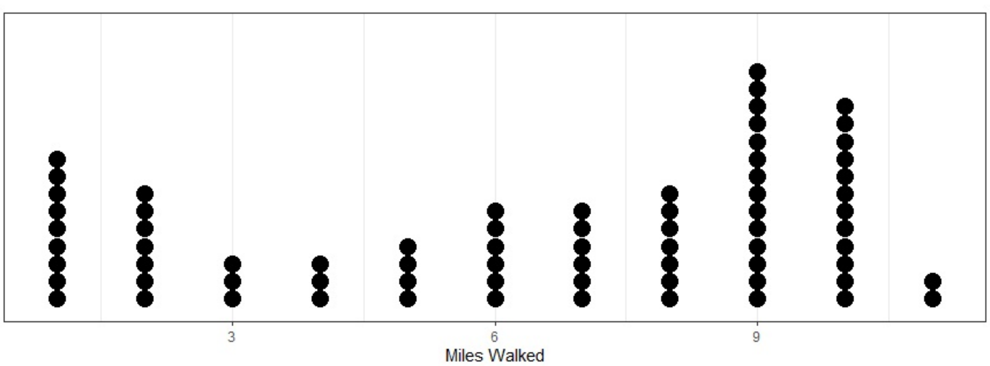
In the first example of distributing cookies, we are computing the arithmetic mean that corresponds with even distribution of the quantity.

Definition 14.1. The **arithmetic mean** of values $a_1, a_2, \dots a_n$ is defined by the formula:

$$\mu = \frac{a_1 + a_2 + \cdots a_n}{n}.$$

In the cookie example we take the total number of cookies (120) and divide them evenly among the 25 students to have an average of 4.8 cookies per person.

Example 14.2. A group of 73 high school students participate in a walk-a-thon at their school[3]. The number of miles walked is represented in the following dot plot and has an arithmetic mean of 6.38 miles.



While we thought of the mean as an even distribution in the cookie example, in this example it is easier to think of the arithmetic mean as a balance point. If we think of the dot plot as a scale with each of the dots having the same weight, the scale is balanced at the point of 6.38.

When finding an average rate of change, one does not take the arithmetic mean of the values, but instead determines the total values of both quantities in the rate of change and then looks at the ratio. In the case of the motorboat, the total distance is 48 miles over a period of 5 hours, giving an average speed of 9.6 miles per hour.

This average of rates of change over equal intervals can be generalized using the harmonic mean.

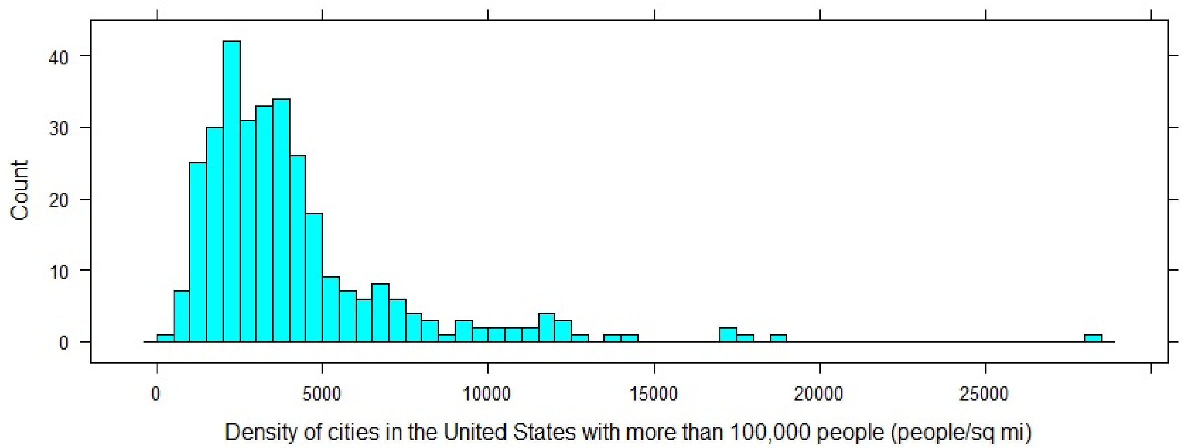
[3]Adapted from https://www.engageny.org/resource/algebra-i-module-2-topic-lesson-1

Definition 14.2. The **harmonic mean** of values $a_1, a_2, \ldots a_n$ is defined by the formula:

$$\frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}}.$$

The harmonic mean is generally used to find the center of data that is a ratio of some type. Such situations include density (mass/volume) in physics, the price earnings ratio (price/earnings) in finance, and fuel economy (miles/gallon) for vehicles.

Example 14.3. When comparing different cities in the United States, it is sometimes useful to not just look at their overall populations, but to look at how many people there are per square mile. This gives a better impression of the density of the city, which usually corresponds to how big a city really feels.



We see from the histogram above that New York City is an outlier in the data set. While the arithmetic mean is 4,171 people/sq mi, the harmonic mean is 2,641 people/sq mi. As we can see from the data, and the context of the variable, this harmonic mean is a better representation of the average density of these cities.

We now turn our attention to the example involving increases and decreases by a certain percentage. For this we are wanting to know what the equivalent percentage increase or decrease would be if it was constant over the four years. We see that the final amount of the painting can be found as

$$1.05(0.90(1.20(1.50(5,000)))) = (1.05 \cdot 0.90 \cdot 1.20 \cdot 1.50)\, 5,000.$$

So the equivalent amount of increase would be a 14% increase, since $\sqrt[4]{(1.05 \cdot 0.90 \cdot 1.20 \cdot 1.50)} = 1.14$. This average of values is called the geometric mean.

Definition 14.3. The **geometric mean** of values $a_1, a_2, \ldots a_n$ is defined by the formula:

$$\sqrt[n]{a_1 a_2 \cdots a_n}.$$

While the most common uses of the geometric mean involve compounded interest, the Water Quality Index produced by the EPA uses a geometric mean to combine multiple water quality indexes into a single index[4]. Using the geometric mean, rather than the arithmetic mean keeps individual extreme values on the sub-indexes from having a large effect on the overall index.

An additional common measure of the 'center' of a data set is the median.

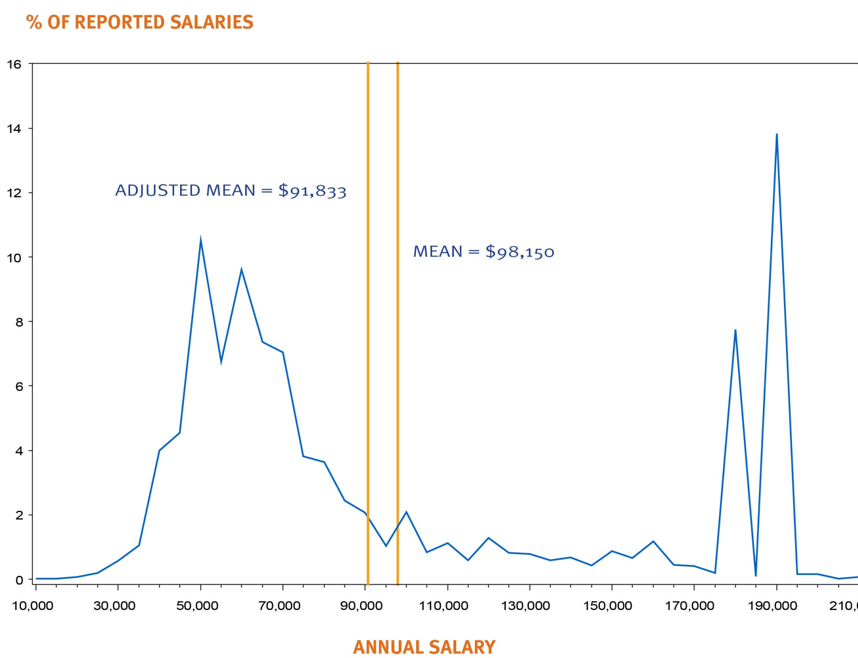
[4]https://19january2017snapshot.epa.gov/sites/production/files/2014-12/documents/water_quality_index_aggregation.pdf

Definition 14.4. The **median** is the value separating the higher half of a data sample, a population, or a probability distribution from the lower half.

The median of a sample is appropriate when the question about the data regards an 'average case', rather than an 'average of cases'. We also generally use the median when the data set has extreme values on one end of the distribution, as these extreme values have a large effect on the arithmetic mean. For these reasons, variables such as salaries and home prices are usually described with medians.

Another time when a mean cannot be used, but a median can, is with ordinal variables. When there is not a set difference between consecutive values of a variable, the means of the variable do not have a good meaning. However, the median still makes sense. For instance, if we want to know how much education the average person in a sample has, we can sort the sample by the amount of education in number of years or type and then identify the educational level of the person in the middle of the distribution.

Example 14.4. There are some distributions for which the center is not a valuable piece of information. Consider the following graph of the starting salaries for lawyers in 2018[5]. The adjusted mean takes into account the expected under reporting of salaries by smaller companies.



For this data set the adjusted mean of $91,833 is likely not much different from the median. However, there are not many lawyers fresh out of law school making that as their starting salaries. What are some other ways that the starting salaries of lawyers should be reported so that people can make a more informed decision about whether going to law school is a good choice for them?

14.4.1 Exercises

1. Is there a difference between the following two questions? (Justify your answer)

    a. What is the average height of students in 6th grade?
    b. What is the height of the average student in 6th grade?
    c. How do the answers to these questions differ when accounting for gender?

    Use Census at School data to provide answers to these questions for a sample of 500 6th grade students.

[5] https://www.nalp.org/salarydistrib

2. Download the current salaries of all players in the NFL.
    a. What is the average salary of NFL players?
    b. What is the salary of the average NFL player?
    c. Why are these numbers similar or different? When would someone report one instead of the other?
    d. How do these numbers compare to other professional sports?
3. Is there a relationship in terms of inequalities for any of the "averages" described in this section? (Test for relationships with 2 numbers in the data set.)

14.5 Measures of Variability

While the center of a data set gives us a partial description of the values of a variable, we need ways to describe how much variability exists in the data to more fully understand the situation.

One aspect of a location that people consider when choosing places to live are how hot it will be. The average high temperature in 2019 in Phoenix and Honolulu were both 86° F[6]. Since the average temperatures are a balance between the highs and lows, these two cities can have the same average temperatures, even though they have drastically different climates. So it is helpful to have numerical measures of how much variation exists.

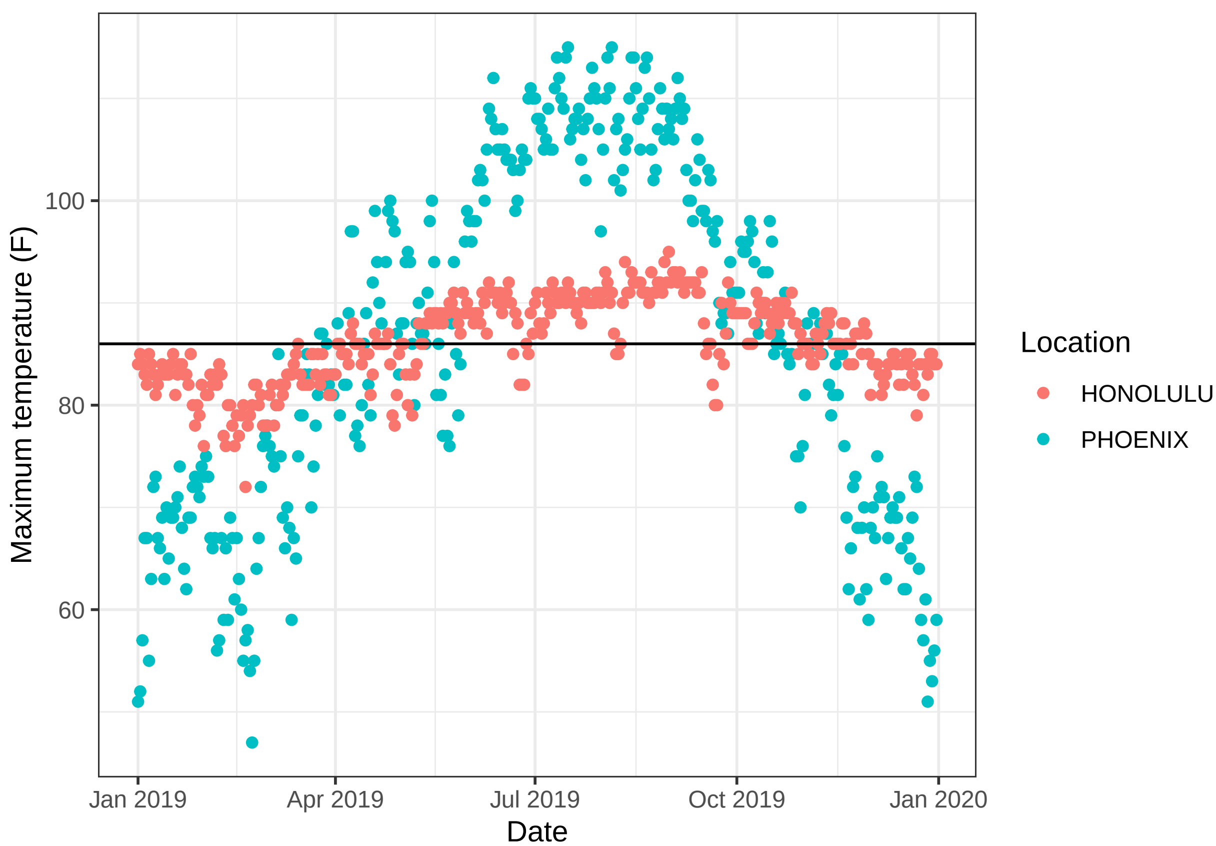


Figure 14.12: Temperatures for Phoenix and Honolulu

The first measure of variability focuses on the range of the middle half of the data.

[6]https://www.ncdc.noaa.gov/cdo-web/

Table 14.1: Five Point Summary of Temperatures for Phoenix and Honolulu

	Q0 (Min)	Q1	Q2 (Median)	Q3	Q4 (Max)	IQR
Phoenix	47	72	87	102	115	28
Honolulu	72	83	86	90	95	7

Definition 14.5. The **innerquartile range (IQR)** is based on the division of the data set into quartiles, Q1 is the median of the lower half of the ranked data set, Q2 is the median of the data set, and Q3 is the median of the upper half of the ranked data set. The IQR of the data set is then Q3-Q1.

In our example comparing the temperatures of Phoenix and Honolulu we see these quartile values are very different.

These values are represented graphically in the box plots with Q1 being the left side of the box, Q3 being the right side of the box, and IQR being the width of the box.

```
ggplot(HonoluluPhoenixClimate, aes(TMAX)) + geom_boxplot(aes(colour = factor(NAME))) + labs(x= "High Te
```

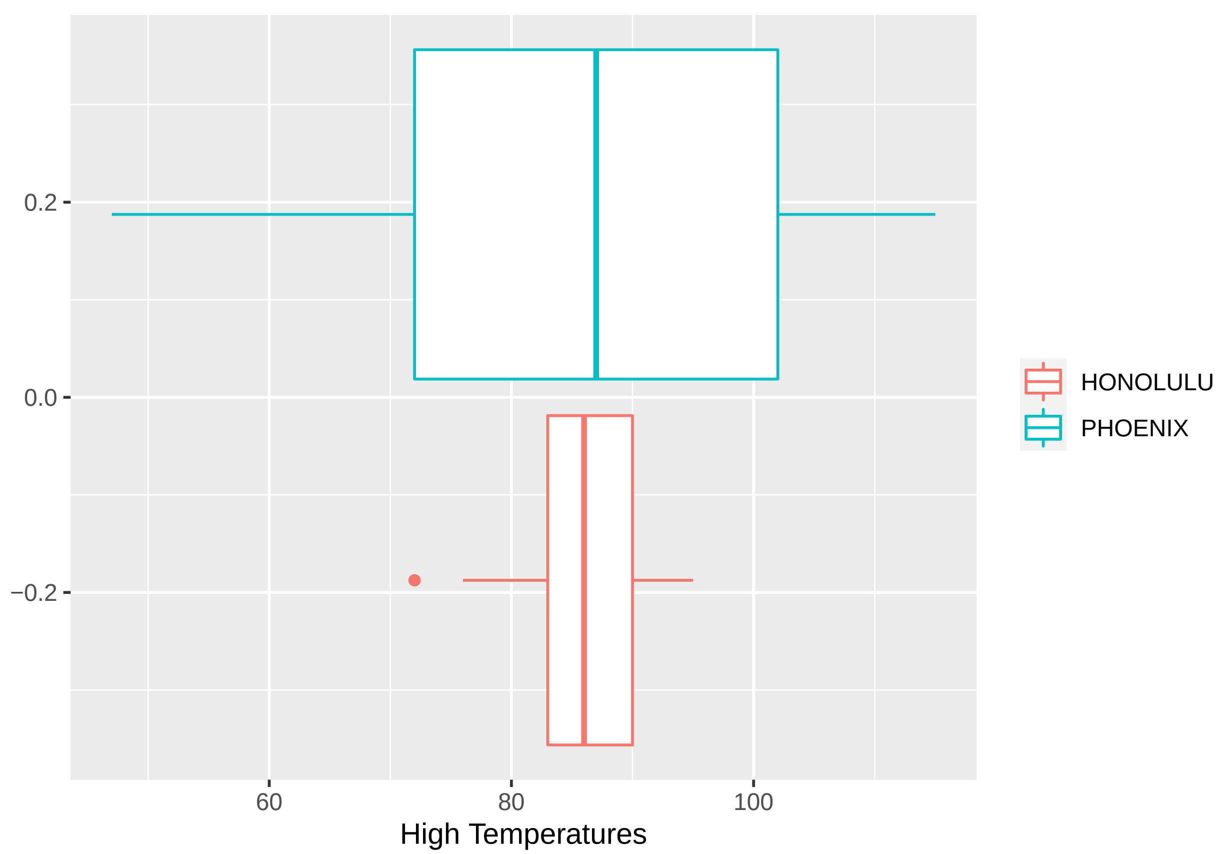


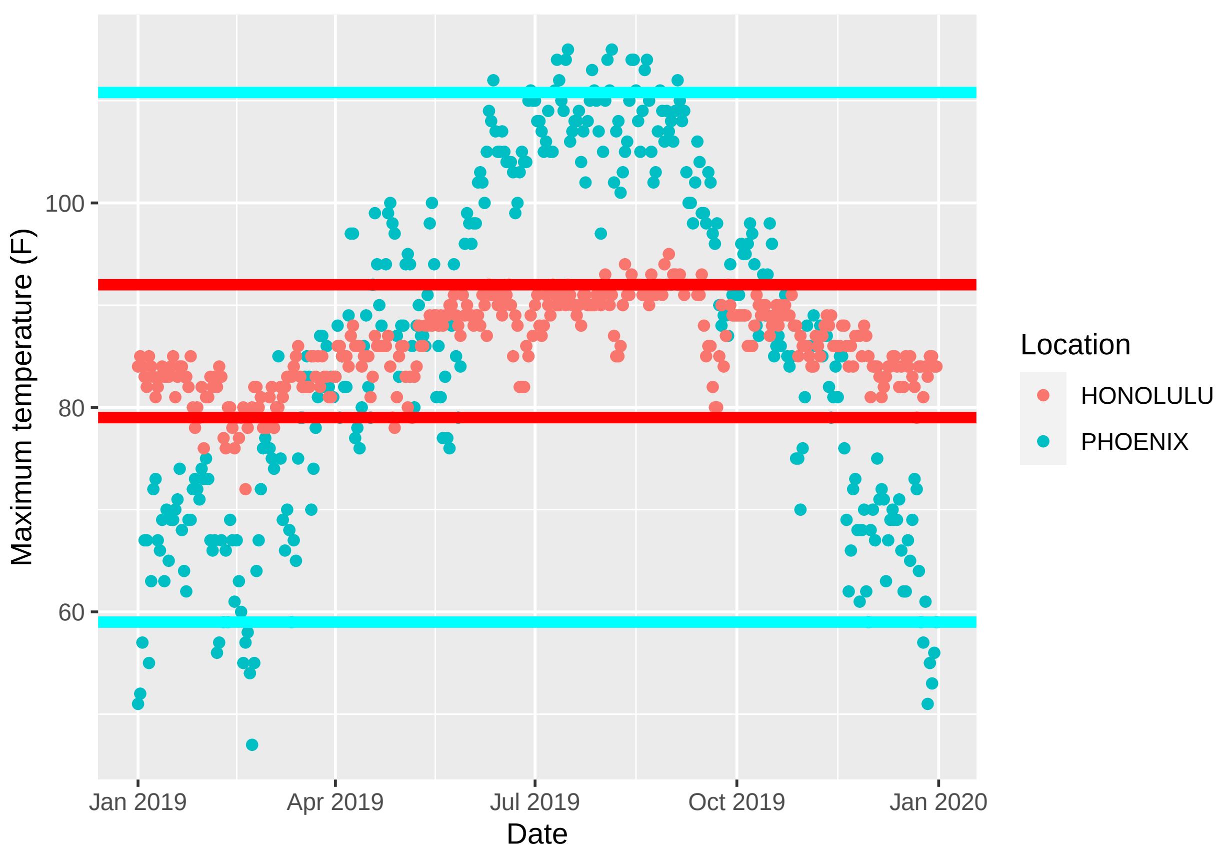
Figure 14.13: Box Plot Comparing Temperatures of Phoenix and Honolulu

While the quartiles are commonly used to describe the variation of a data set in terms of its middle values, we will see in later chapters that with hypothesis testing it is standard to look for the location of 95% of the values in a data set.

Definition 14.6. A **95-percent coverage interval** describes the interval within which 95% of the data exists in a data set.

For the temperature example, we find that 95% of the high temperatures in Phoenix lie in the interval $(59.0, 110.8)$, while in Honolulu 95% of the high temperatures lie in the interval of $(79.0, 92.0)$. This further differentiates the climate of Honolulu from Phoenix in that almost every day in Honolulu is within most people's comfort level, while there are many days in Phoenix for which people are uncomfortable being outside.

```
ggplot(HonoluluPhoenixClimate, aes(DATE, TMAX)) + geom_point(aes(colour = factor(NAME))) + labs(x="Date
```



Each of the prior methods of measuring variability involves ranking the data and then describing the data that fits within certain intervals. Because these measures of variability depend only on the rank of the data, they can be used with either ordinal or continuous data.

When a variable is a continuous numeric variable, another way to describe the variability is to measure how much the data set differs from one of the centers of the data set. With such measurements, we want the value to become more reliable as we increase the number of points in our data set and so we want to measure some type of average of distances from the center. The most basic of these measure is the mean absolute deviation.

Definition 14.7. The **mean absolute deviation** of a data set is the mean of the absolute values of the differences from the center of the data set. If our data set is $\{x_1, x_2, x_3, \ldots, x_n\}$, with center of $\overline{x}$, then the mean absolute deviation of the set is

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - \overline{x}|.$$

While any center of the data set could be used, the most common center for the mean absolute deviation is the arithmetic mean.

The mean absolute deviation can be found using spreadsheets with the function `AVEDEV` and in R using `MeanAD`.

For Phoenix, the mean absolute deviation in the temperatures is 14.5 degrees, while Honolulu is only 3.4 degrees.

Related Content Standards

- (6.SPB.5) Summarize numerical data sets in relation to their context, such as by:
  c. Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.
  d. Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

While the mean absolute deviation is easy to compute and interpret, it does not work well with mathematical statistics based on the theory of probability. So the most common measures of variability are based on the average of the squares of the differences from the mean, similar to the distance formula between points in $\mathbb{R}^n$.

Definition 14.8. The **standard deviation** of a data set $\{x_1, x_2, \ldots, x_n\}$ with mean of $\overline{x}$ is given by

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

The **variance** of the data set is the square of the standard deviation,

$$\mathrm{Var} = \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

We will look at some of the key properties of the variance in the next chapter on probability, but one key aspect of the variance is that it is equal to the difference between the mean of the squares and the square of the means of the values in the data set,

$$\sigma^2 = \left(\frac{1}{n}\sum_{i=1}^{n}x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2.$$

The spreadsheet functions to find the standard deviation and variance for a population are `STDEV` and `VAR` with the corresponding functions in R being `sd` and `var`. If the data set represents a sample of a larger population, the formulas and commands are slightly different and will be discussed in Chapter 15.

When we compute the standard deviation and the variance of the high temperatures for Phoenix and Honolulu we find that the standard deviations (16.98 and 4.10 degrees, respectively) are similar to the mean absolute deviations, with the standard deviations being slightly larger.

14.5.1 Exercises

1. Use the Census at School Random Sampler https://ww2.amstat.org/censusatschool/ to create a spreadsheet with 1000 random students. Use this data set to explore the variability of the variables of `Age_years`, `Height_cm`, `Travel_time_to_school`, and `Armspan_cm` using the techniques discussed in this section. Create a report to discuss your findings.

14.6 Exploring Bivariate Data Numerically

14.6.1 Categorical × Numerical Data

One of the primary uses of numerical statistics measuring the center and variation of data sets is to compare the properties of two or more data sets based on a categorical variable. For instance, if we look at the data from the 1985 Current Population Survey, `CPS85` in the MosaicData R package we can compare wages (in dollars per hour) for different genders.

We will first look at the possible centers that we can use to compare the populations.

Centers for Wages by Gender

Median

Arithmetic Mean

Harmonic Mean

Geometric Mean

Male

$8.93

$9.99

$7.52

$8.72

Female

$6.80

$7.88

$6.18

$6.92

In order to help us to determine which of these centers to use, we need to consider the properties of the variable that we are studying. Since a wage is a rate (dollars per hour), so the harmonic mean would be the most appropriate of the means to use. However, one could also easily justify the use of the median to say that the average female has a lower wage than the average male.

Related Content Standards

- (7.SPB.4) Use measures of center and measures of variability for numerical data from random samples to draw informal comparative inferences about two populations.
- (HSS.ID.2) Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets.
- (HSS.ID.3) Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

This choice of center is further justified when looking at the density plot of the wages for the 534 individuals sampled and noticing that it is skewed right, with some significant outliers, and has a distribution that resembles a lognormal distribution (for which the harmonic mean is considered the appropriate center).

```
library(mosaicData)
ggplot(CPS85, aes(x=wage)) + geom_density(kernel = "gaussian", aes(colour = factor(sex))) + labs(x= "Wa
```

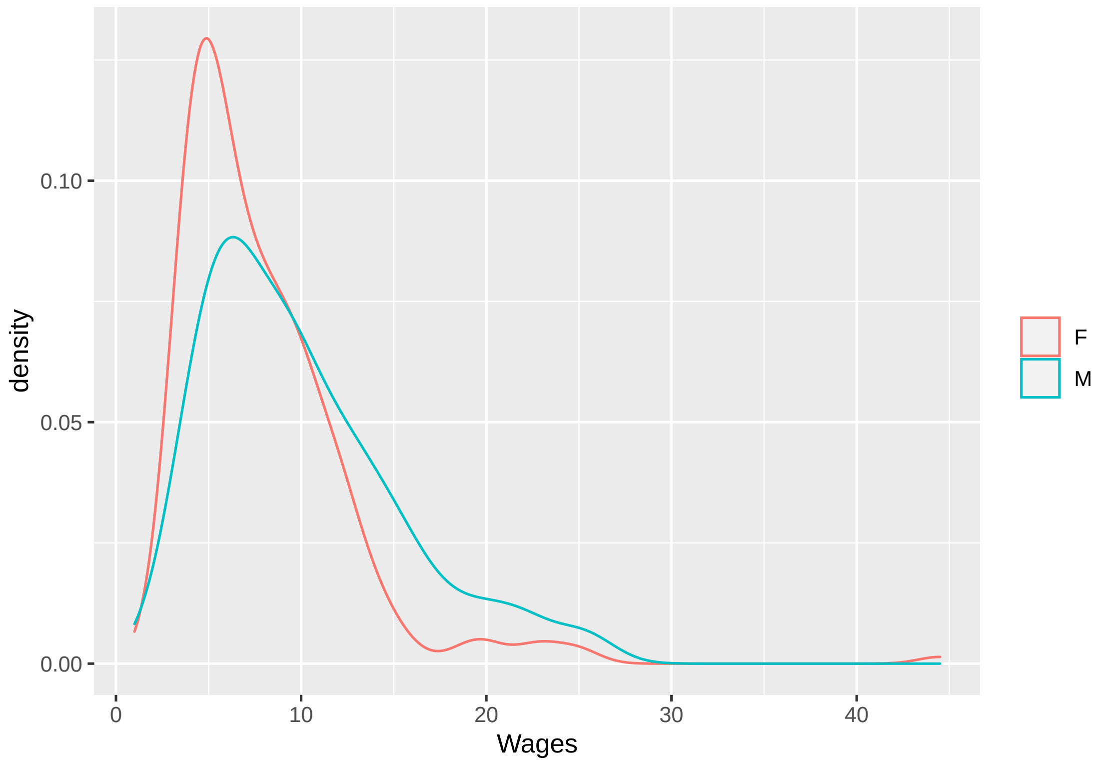


Figure 14.14: Density Plot of Wages

Table 14.2: Five Point Summary of Wages

	Q0 (Min)	Q1	Q2 (Median)	Q3	Q4 (Max)	IQR
Male	\$1.00	\$6.00	\$8.93	\$13.00	\$26.29	\$7.00
Female	\$1.75	\$4.75	\$6.80	\$10.00	\$44.50	\$5.25

We can also use the quartiles of the two distributions and the inner quartile range, along with the corresponding box plots, to further clarify the differences between male and female wages.

We notice that there is less variability in the female wages and 75% of the females in the sample make less than \$10.00 per hour, where just under half of the males make more than \$10.00 per hour.

14.6.2 Numerical × Numerical Data

When we are exploring the relationship between two variables that are numerical, we would like to find some type of numerical description of the strength of association between the variables. If the numerical variables are both either ordinal or continuous so that the data set for each variable can be ranked in order, then the Kendall rank correlation coefficient (Kendall's τ) can be used to measure how well the relationship between the two variables can be described using a monotonic function.

A pair of observations (x_i, y_i) and (x_j, y_j) are called concordant if $(x_j - x_i)$ and $(y_j - y_i)$ have the same sign. The pair is disconcordant if they have opposite signs. If either of the differences are zero, the pair is said to be tied.

Definition 14.9. For a set of data points, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where the x_i and y_i correspond with ordinal or continuous variables without any tied pairs, **Kendall's correlation coefficient** is given by

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\binom{n}{2}},$$

where $\binom{n}{2} = \frac{n(n-1)}{2}$ is the number of ways to choose pairs from n items.

If the data set include tied pairs, an adjustment in the formula is needed to create measures τ_b or τ_c, which are the standard outputs from statistical applications.

When the variables are continuous, fit a normal distribution, and have a linear relationship, the strength of the association between the variables is given by the Pearson correlation coefficient (Pearson's ρ).

Definition 14.10. For a set of data points, $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where the x_i and y_i correspond with continuous variables, **Pearson's correlation coefficient** is given by

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}},$$

where $\overline{x}$ and $\overline{y}$ are the arithmetic means of the two data sets.

We can explore these two measures of association using the nutritional information from McDonald's[7].

We see that if the relationship is close to a linear relationship that both ρ and τ are close to 1. We can also see that for these variables, both ρ and τ generate similar results. If instead, the relationship is strong, but not linear, then the two correlation coefficients differ.

Related Content Standards

- (8.SPA.2) Know that straight lines are widely used to model relationships between two quantitative variables. For scatter plots that suggest a linear association, informally fit a straight line, and informally assess the model fit by judging the closeness of the data points to the line.
- (8.SPA.3) Use the equation of a linear model to solve problems in the context of bivariate measurement data, interpreting the slope and intercept.

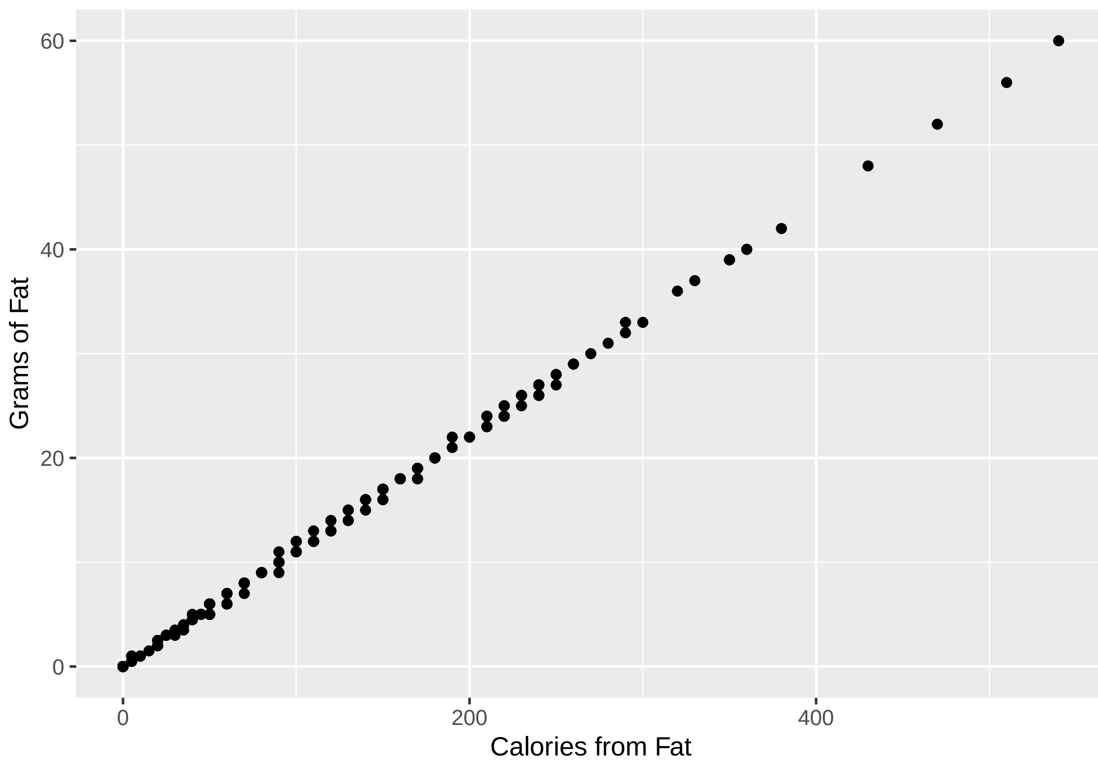


Figure 14.15: Fat and Calories from Fat (Pearson: 0.9996, Kendall: 0.9942)

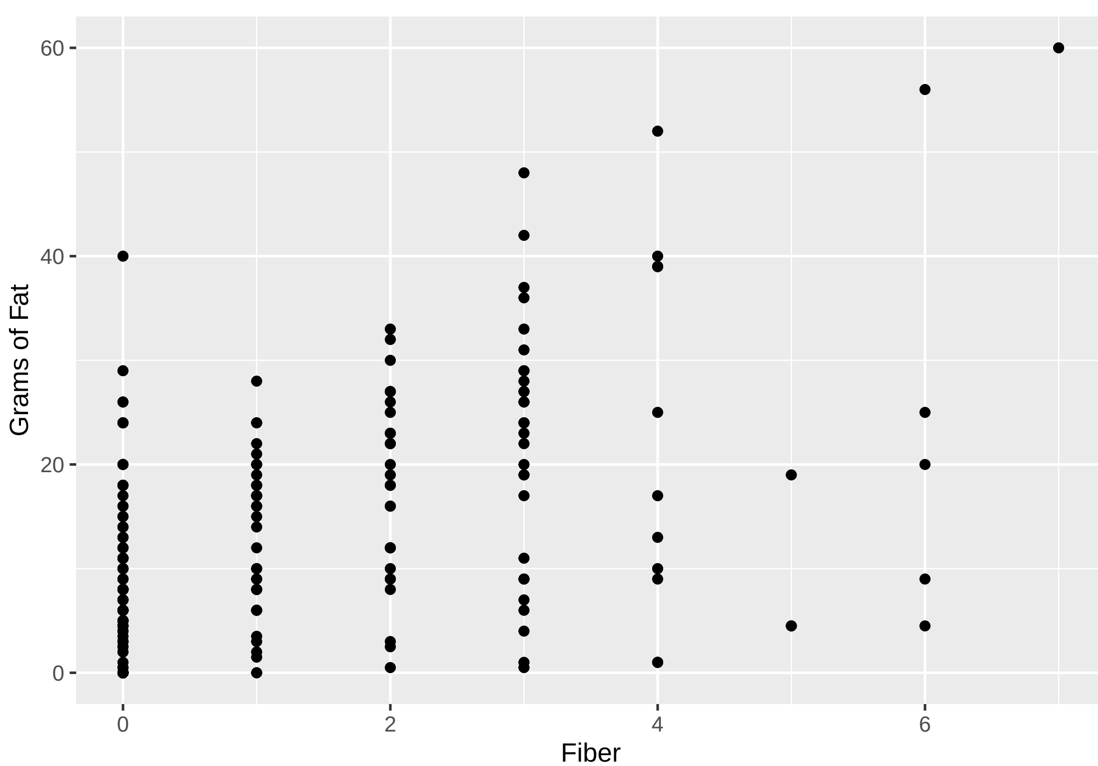


Figure 14.16: Fiber and Grams of Fat (Pearson: 0.5739, Kendall: 0.4679)

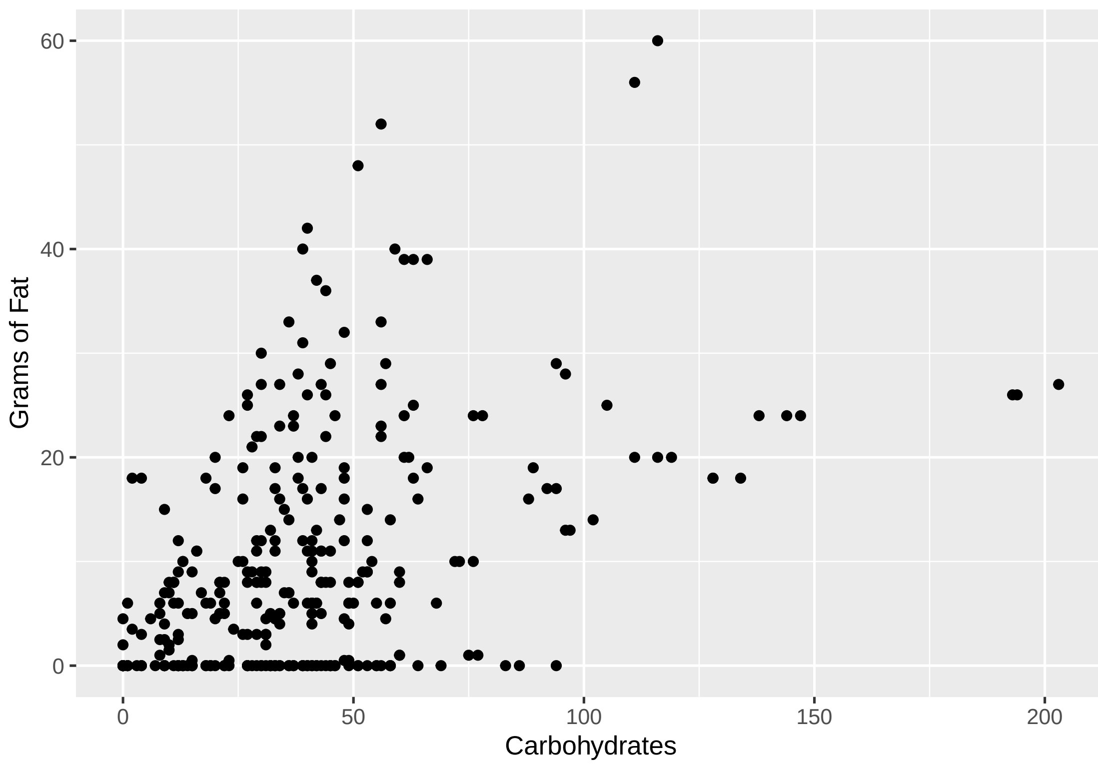


Figure 14.17: Carbohydrates and Grams of Fat (Pearson: 0.4422, Kendall: 0.2985)

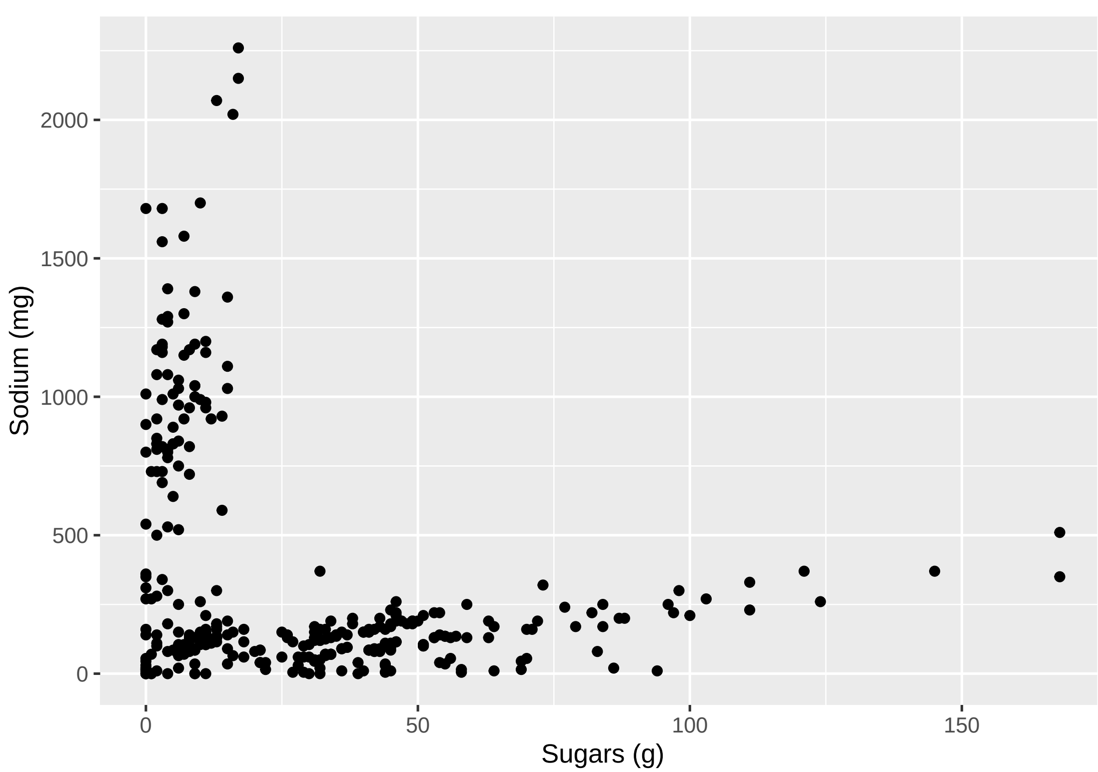


Figure 14.18: Sugar and Sodium (Pearson: -0.2935, Kendall: -0.0895)

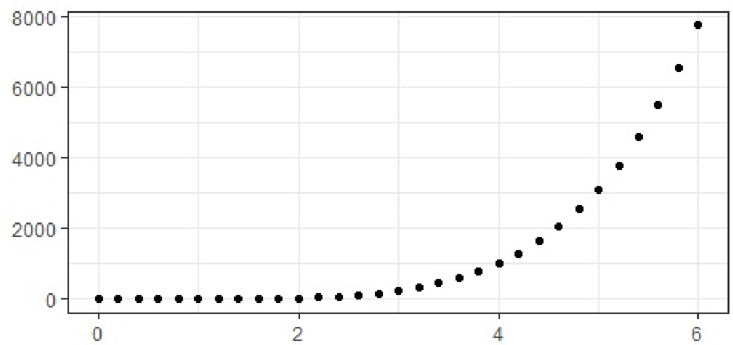


Figure 14.19: Example with Pearson correlation of 0.8171 and Kendall correlation of 1

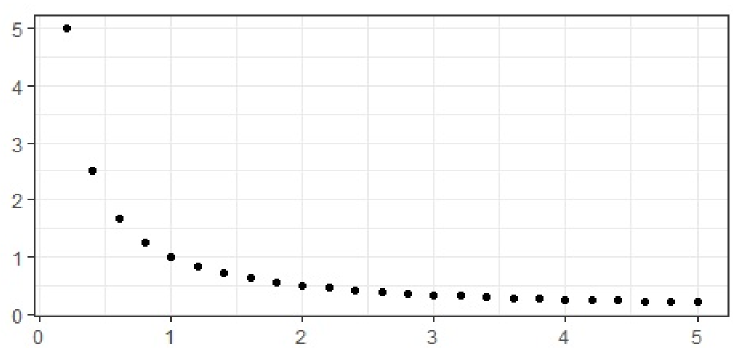


Figure 14.20: Example with Pearson correlation of -0.6747 and Kendall correlation of -1

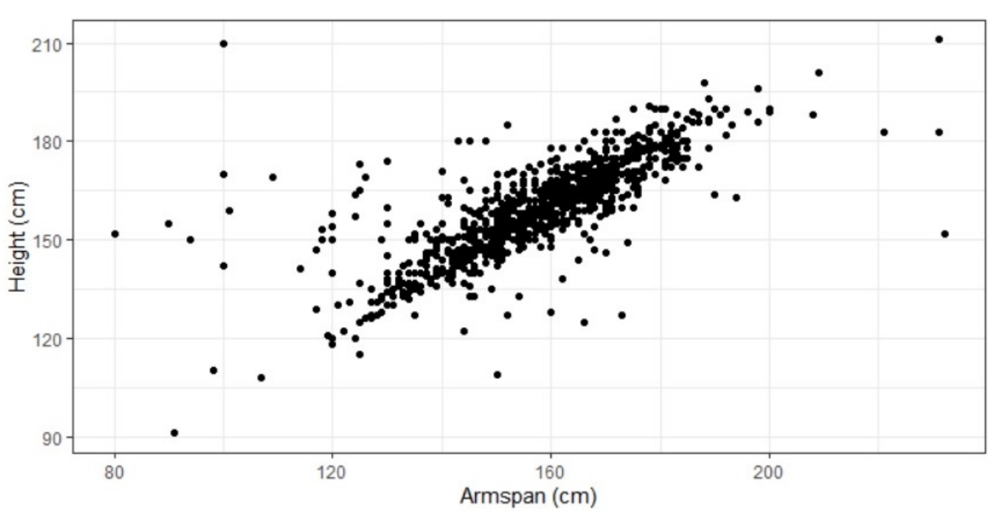


Figure 14.21: Height and Armspan

Looking at a population of 1000 students from around the world using the Census at School random sampler, we can look at how the height and armspan of the individuals in this population may be related.

Looking at the scatterplot of the data, it appears that there may be a linear relationship between the two variables. Because of the possible linearity of this relationship, we look at the Pearson's correlation coefficient and see that $\rho = 0.7627$, confirming a possible linear relationship.

We would now like to estimate the linear relationship. In doing so, we are looking for the equation of a line that 'best' fits the data. If we let y represent a person's height (in cm) and x represent a person's armspan (in cm), we model the relationship with a function of the form $y = mx + b$.

One method of measuring how well the linear model fits the data is to measure the difference between the actual y-value and the predicted y-value for each data point. We call this value the **residual**.

$$\text{Residual} = \text{actual } y\text{-value} - \text{predicted } y\text{-value}$$

If we write our data points as ordered pairs, $\{(x_i, y_i)\}$, we let

$$r_i = y_i - (m \cdot x_i + b)\,.$$

One idea is to find the values for m and b that minimize the sum of these residuals. However, a large positive residual in one value could compensate for a large negative residual in another. So one way that we could compensate for this would be to minimize the sum of the absolute values of the residuals. While this option does work, to find the Least Absolute Deviations (LAD) regression, it is not the most common. Instead, the common method is to minimize the sum of the squares of the residuals,

$$S = \sum_{i=1}^{n} r_i^2,$$

to determine the **least squares approximation**.

If we let

$$S(m,b) = \sum_{i=1}^{n} (y_i - m \cdot x_i - b)^2$$

we can minimize this function by looking at its partial derivatives,

$$\frac{\partial S}{\partial m} = \sum_{i=1}^{n} (2 \cdot (y_i - m \cdot x_i - b) \cdot -x_i) = \left(\sum_{i=1}^{n} -2x_i y_i\right) + \left(\sum_{i=1}^{n} 2x_i^2\right) m + \left(\sum_{i=1}^{n} 2x_i\right) b$$

and

$$\frac{\partial S}{\partial b} = \sum_{i=1}^{n} (2 \cdot (y_i - m \cdot x_i - b) \cdot (-1)) = \left(\sum_{i=1}^{n} -2y_i\right) + \left(\sum_{i=1}^{n} 2x_i\right) m + \left(\sum_{i=1}^{n} 2\right) b$$

We then find the minimal value of S where these derivatives are both zero, which is the solution to the system of equations

$$\left(\sum_{i=1}^{n} 2x_i\right) m + \left(\sum_{i=1}^{n} 2\right) b = \sum_{i=1}^{n} 2y_i \tag{14.1}$$

$$\left(\sum_{i=1}^{n} 2x_i^2\right) m + \left(\sum_{i=1}^{n} 2x_i\right) b = \sum_{i=1}^{n} 2x_i y_i \tag{14.2}$$

[7]http://nutrition.mcdonalds.com/nutrition1/nutritionfacts.pdf

We can see from the first equation in the system that this is equivalent to

$$2n \cdot \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) m + 2nb = 2n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$$

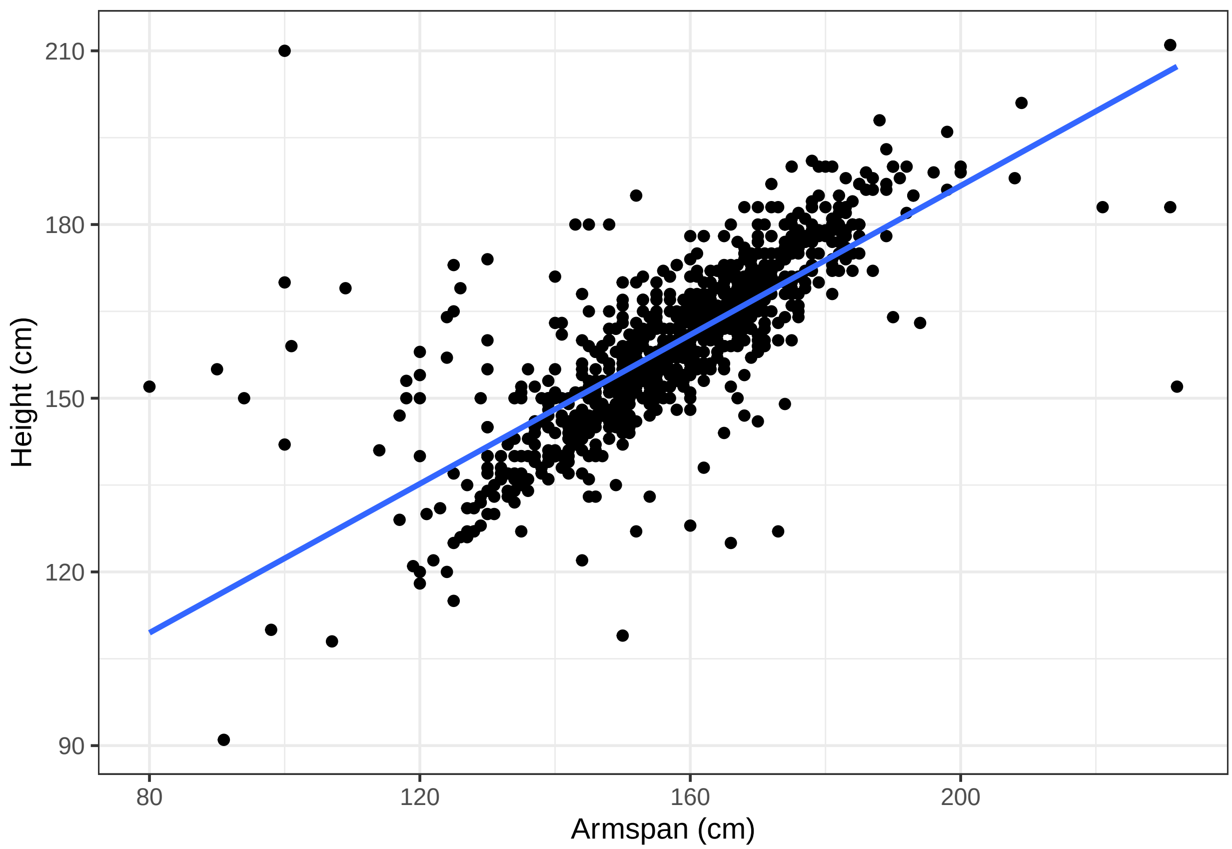
which can be written as

$$\overline{y} = m\overline{x} + b$$

where $\overline{x}$ and $\overline{y}$ are the arithmetic means of the x_i and y_i, respectively. This implies that the point $(\overline{x}, \overline{y})$ always lies on the least squares linear approximation.

For the example of the height and the armspan, the least squares linear approximation is that

$$\text{Height (cm)} = 0.64 \cdot \text{Armspan (cm)} + 58.01$$



This means that for the population that we are analyzing that an increase of 1 cm in the armspan of a person would correspond to an increase of 0.64 cm in the height of the person. For this approximation, we have that the residual

In order to determine how well this line fits the data, we want to compare the total sum of squares

$$\sum_{i=1}^{n}(y_i - \overline{y})^2$$

to the residual sum of squares

$$\sum_{i=1}^{n}(y_i - (m \cdot x_i + b))^2$$

using

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - (m \cdot x_i + b))^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}.$$

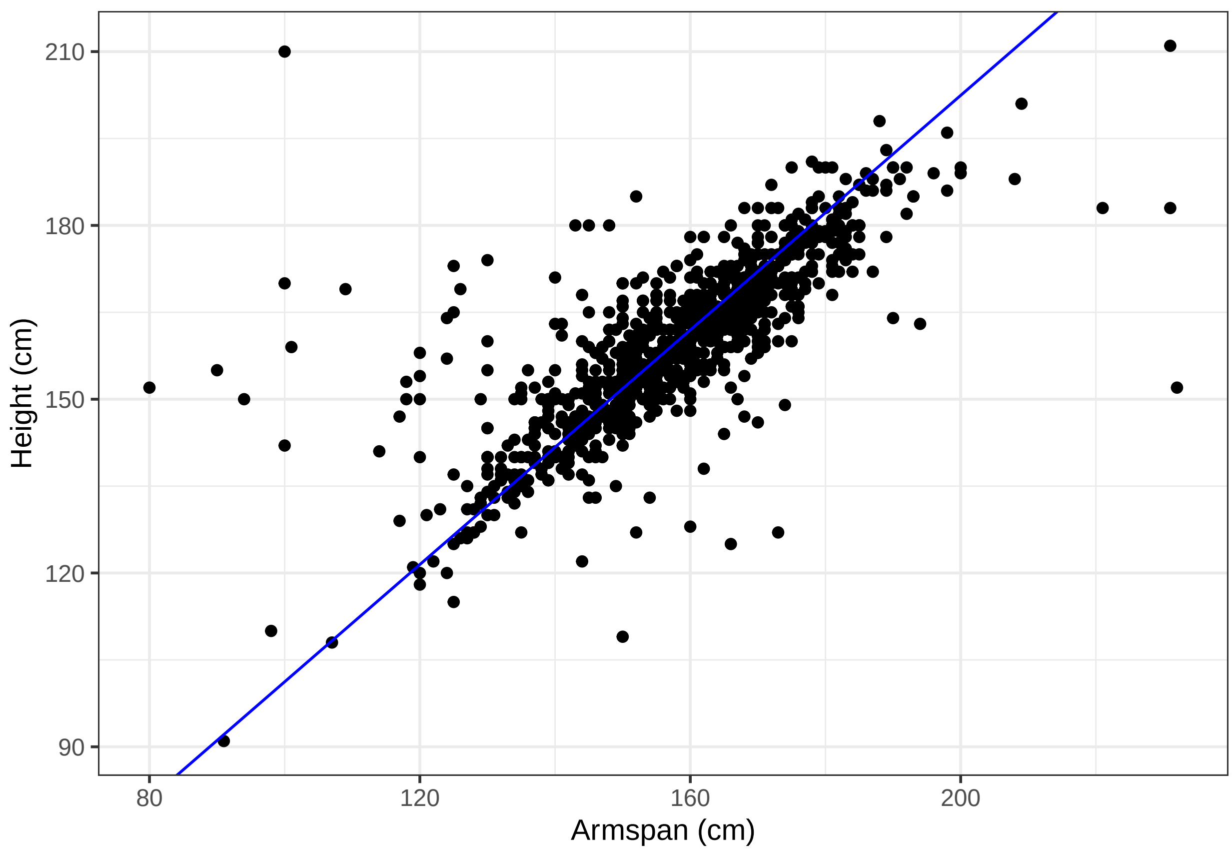
This parameter is the proportion of the variance in the dependent variable (y) that is predictable from the independent variable (x).

For the example of height and armspan, $R^2 = 0.58$ meaning that 58% of the variance in the height can be explained by the armspan.

Looking at the scatterplot for armspan and height, we notice that there is a large cluster in the middle of the data that seems to have a different slope than the estimated line. We also notice that the model given by the least squares approximation says that a person without any armspan would still be 58 cm tall. Using this information, we could insist that the intercept for the linear approximation would be at the origin with the interpretation that someone with zero armspan would also have zero height.

With this additional constraint, we can see from the sum of squares of the residual that $m = \frac{\overline{y}}{\overline{x}}$. So the data would be modeled by

$$\text{Height (cm)} = 1.01 \cdot \text{Armspan (cm)}$$



So we see that even though this does not minimize the sum of squares, it may be the best approximation for our data in terms of the portion of the population for which we are focused and within the context of the situation.

14.6.3 Categorical × Categorical Data

Related Content Standards

Table 14.3: Transportation to School (Frequency)

	Bus	Car	Walk	Other	Total
Australia	63	137	56	20	276
Canada	103	62	77	18	260
New Zealand	39	45	43	9	136
United Kingdom	33	28	36	8	105
United States	50	151	12	10	223
Total	288	423	224	65	1000

Table 14.4: Transportation to School (Conditional Row Relative Frequency)

	Bus	Car	Walk	Other	Total
Australia	22.8%	49.6%	20.3%	7.2%	100%
Canada	39.6%	23.8%	29.6%	6.9%	100%
New Zealand	28.7%	33.1%	31.6%	6.6%	100%
United Kingdom	31.4%	26.7%	34.3%	7.6%	100%
United States	22.4%	67.7%	5.4%	4.5%	100%

- (8.SPA.4) Understand that patterns of association can also be seen in bivariate categorical data by displaying frequencies and relative frequencies in a two-way table. Construct and interpret a two-way table summarizing data on two categorical variables collected from the same subjects. Use relative frequencies calculated for rows or columns to describe possible association between the two variables.
- (HSS.ID.5) Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

When analyzing the interactions between two categorical variables, the first step usually involves the creation of a **two-way frequency table** that organizes the number of cases for each possible combination of values for the two variables. Using a random sample of 1000 students from the Census at School site we can explore possible relationships between the country and the way that the students get to school. In order to simplify the table we combined all of the categories with just a few cases into a general ‘other’ category.

Just looking at the counts by themselves can sometimes be valuable, but can also lead to poor conclusions. For instance, we may notice that 28 students in the UK travel to school by car, while 62 students in Canada also travel to school by car. Someone just looking at these raw counts may think that students are twice as likely to travel to school by car in Canada as they are in the UK. However, to make such a statement we should instead use the **conditional relative frequency tables** that are conditioned on each of the two variables.

If we condition on the country, we see that a higher percentage of the students in the sample from the U.K., rather than Canada, travel by car to school.

Looking at this conditional relative frequency table we can also more easily notice that the students from the United States in the sample are much less likely to walk to school than those from the other countries.

If we condition on the method of transportation to school, we see that of the people in the sample that take a car to school, 35% of them are from the United States, while only 6.6% of the car riders are from the United Kingdom.

One might also use a **joint relative frequency table** where each of the entries represents the proportion of the overall sample being analyzed.

Table 14.5: Transportation to School (Conditional Column Relative Frequency)

	Bus	Car	Walk	Other
Australia	21.9%	32.4%	25%	30.8%
Canada	35.8%	14.7%	34.4%	27.7%
New Zealand	13.5%	10.6%	19.2%	13.8%
United Kingdom	11.5%	6.6%	16.1%	12.3%
United States	17.4%	35.7%	5.4%	15.4%
Total	100%	100%	100%	100%

Table 14.6: Transportation to School (Joint Relative Frequency)

	Bus	Car	Walk	Other	Total
Australia	6.3%	13.7%	5.6%	2.0%	27.6%
Canada	10.3%	6.2%	7.7%	1.8%	26.0%
New Zealand	3.9%	4.5%	4.3%	0.9%	13.6%
United Kingdom	3.3%	2.8%	3.6%	0.8%	10.5%
United States	5.0%	15.1%	1.2%	1.0%	22.3%
Total	28.8%	42.3%	22.4%	6.5%	100%

14.6.4 Exercises

1. How do the salaries of NFL players, NBA players, and MLB players compare?
2. Below are test results for eighth graders from four local middle schools based on the proficiency levels of the exam for 2018-2019.

a. How could you use this information to determine where to buy a house based on where you would send your child to middle school? Would it change based on the characteristics of your child?
b. Would any of your thoughts change based on the following demographic information about these schools?

3. The following table describes the age distributions of people in the United States and Japan. Compare and contrast the different populations. (from CIA Factbook, downloaded 8/28/2020)

4. Below is the population of the United States from the Census of since 1940. Create a linear approximation for the population with the independent variable being years since 1940. Determine how well the linear approximation fits the data.

Table 14.7: School Test Results

	Level 1	Level 2	Level 3	Level 4
EMS	18.88%	22.89%	29.72%	28.51%
CRMS	40.00%	34.86%	16.00%	9.14%
NMS	21.96%	16.86%	25.88%	35.29%
TMS	1.92%	3.85%	21.15%	73.08%

Table 14.8: School Characteristics

	English Learners	Students with Disabilities	Econom. Disadv.
EMS	5.5%	13.02%	38.55%
CRMS	8.1%	18.38%	69.17%
NMS	1.99%	11.58%	33.75%
TMS	0%	0.61%	28.66%

Table 14.9: Population Demographics

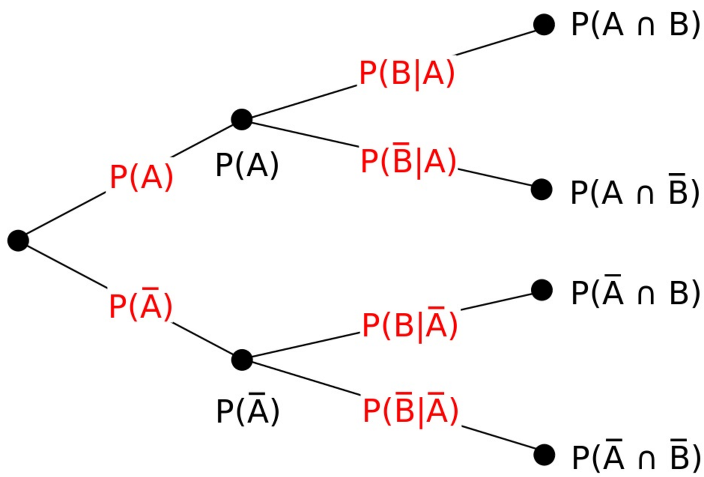
	U.S. Males	U.S. Females	Japan Males	Japan Females
0-14 years	31,374,555	30,034,371	8,047,189	7,623,767
15-24 years	21,931,368	21,006,463	6,254,352	5,635,377
25-54 years	64,893,670	64,565,565	22,867,385	23,317,140
55-64 years	20,690,736	22,091,808	7,564,067	7,570,732
65 years and over	25,014,147	31,037,419	16,034,973	20,592,496

Table 14.10: U.S. Population by Census

	Population
1940	132,164,569
1950	151,325,798
1960	179,323,175
1970	203,211,926
1980	226,545,805
1990	248,709,873
2000	281,421,906
2010	308,745,538
2020	331,449,281

Chapter 15

Samples, Simulations, and Probability



15.1 Probability Overview

Recall the three main foundations of data analysis of trying to understand situations in their context, justification of claims using statistical techniques, and looking at the world through the lens of variability being everywhere. **Probability** is the study of uncertainty and is the foundation for understanding random variability. A phenomenon is considered to be **random** if there are multiple potential outcomes and there is uncertainty about which outcome will occur.

Phenomena like flipping a coin, drawing a ball from a bingo machine, or dealing a shuffled deck of cards are examples of physical randomness. We also have **random selection** in studies, which involves selecting a

Table 15.1: Distribution of Proportion of Coins

	Number of Heads	Number of Tails	Proportion of Heads
3 Trials	3	0	1.000
5 Trials	3	2	0.6000
10 Trials	4	6	0.4000
100 Trials	52	48	0.5200
1,000 Trials	492	508	0.4920
10,000 Trials	4,995	5,005	0.4995

sample of individual cases at random from a population. Some examples are political polls to determine who people prefer in an election or biological samples from a stream to determine the overall health of the water source. Another use of randomness involves the **random assignment** of research participants into different control or experimental groups.

With random phenomena there is uncertainty about individual outcomes, but there is a regular distribution of outcomes with a large enough number of repetitions. For instance, in the table below we give a few examples of simulated outcomes of coin flips for different numbers of repetitions.

We can create such simulations using the `RandBetween(0,1)` function in a spreadsheet the appropriate number of times. Alternatively, we can use the `mosaic` package in R along with the commands `coin <- c(0,1)` and finding the mean of resamples by using `mean(resample(coin,n))` where `n` is replaced by the number of trials.

This sampling technique demonstrates the phenomena that if someone could flip a fair coin an extremely large number of times that half of the time it would land on heads and half of the time it would land on tails.

Just because something is a random phenomena does not mean that all of the outcomes have the same amount of certainty. For instance, in a Presidential election the tickets from the two major parties are much more likely to win the election than someone from a different party. Similarly, not all college football teams start the season with the same likelihood of winning a national championship.

Related Content Standards

- (7.SP.5) Understand that the probability of a chance event is a number between 0 and 1 that expresses the likelihood of the event occurring. Larger numbers indicate greater likelihood. A probability near 0 indicates an unlikely event, a probability around 1/2 indicates an event that is neither unlikely nor likely, and a probability near 1 indicates a likely event.

The **probability** of an event is a number in the interval $[0, 1]$ measuring the likelihood of the event. We often describe these probabilities in terms of percentages with 100% corresponding to certainty and 50% corresponding to the event having the same likelihood of occurring or not occurring.

When there are only a finite number of possible outcomes we define the probability of a certain outcome as

$$P(\text{certain event}) = \frac{\text{number of possible outcomes that correspond to the event}}{\text{number of all possible outcomes}}.$$

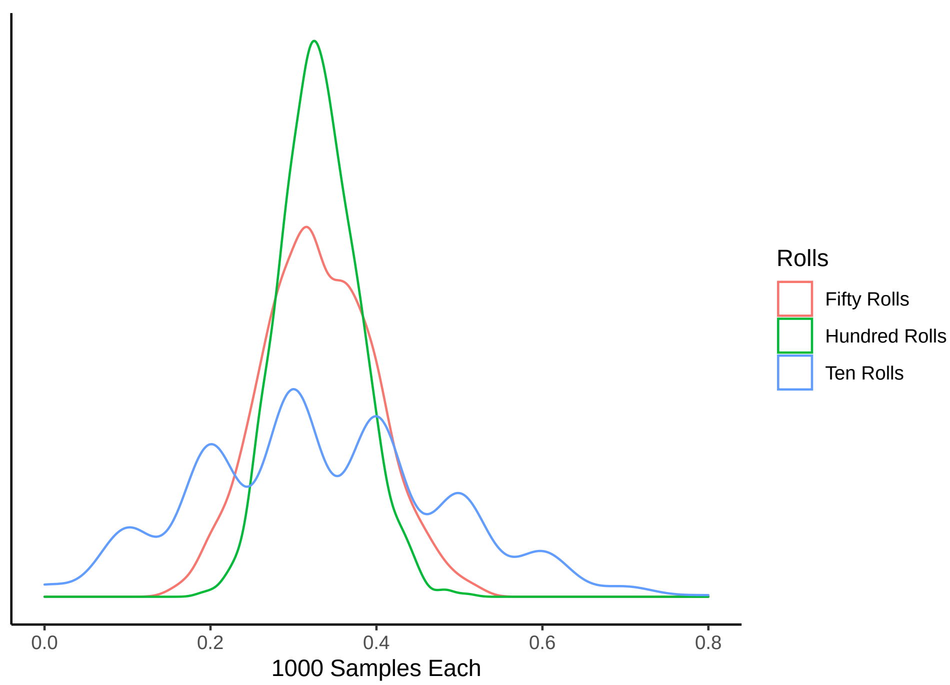
For instance, if we have a fair six-sided die the probability that we roll a number divisible by 3 is $\frac{1}{3}$ because there are two number that are divisible by 3 out of the six possible numbers.

Since we know that there is a probability of $\frac{1}{3}$ of rolling a number divisible by 3 on a fair six-sided die, we would expect that if we collected samples that about a third of the rolls would be a 3 or 6.

Since the **relative frequency** of an event is the number of times that the event occurs during experimental trials divided by the total number of trials conducted, the predicted relative frequency of rolling a 3 or a 6 would be one third.

If we conducted an experiment where we roll a die a certain number of times, the long-run relative frequency of rolling a 3 or 6 should be around one third. The more times one rolls the die in the experiment, the more likely the long-run relative frequency will be closer to one third.

Consider the experiment of rolling a die 10, 50, and 100 times. The density plot below results from a simulation of 1000 cases of this experiment and we can see that the data is centered around $\frac{1}{3}$.



In the case of 10 rolls, there is an point where none of the 10 rolls included a 3 or 6. We also can see that the standard deviation is 1.146. If we increase the number of rolls for each experiment up to 50 rolls, we see that more of the samples are within 0.1 of the expected probability of $\frac{1}{3}$ and the standard deviation is reduced to 0.067. However, out of the 50 experiments there are some that substantially differ from the expected outcome.

As we continue to increase the number of rolls for each experiment up to 100 rolls, the variability in the distribution continues to decrease to the point that the standard deviation drops to 0.046. However, even with this decrease in standard deviation it is still possible to have experiments that differ greatly from the expected probability.

With the rolling of the dice, we can use the theoretical probability to predict the relative frequency and notice that increasing the number of trials greatly increased our accuracy. This phenomena of the average of the long-term relative frequency approaching the theoretical average as the number of cases increases is called the **law of large numbers**.

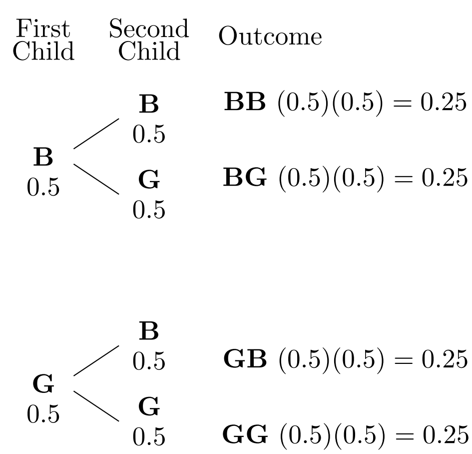
15.1.1 Compound Events

Most of the time that we calculate probabilities the event space is more complicated and is a combination of multiple events. Such events are called compound events. We can use various techniques such as listing out all of the possible events and then work to determine the probabilities for each event, we can use tree diagrams to simplify the process of making sure that we have listed all of the possible events and computed their probabilities, or we can run simulations to estimate the probabilities of the combinations.

Related Content Standards

- (7.SP.8) Find probabilities of compound events using organized lists, tables, tree diagrams, and simulation.
  a. Understand that, just as with simple events, the probability of a compound event is the fraction of outcomes in the sample space for which the compound event occurs.
  b. Represent sample spaces for compound events using methods such as organized lists, tables and tree diagrams. For an event described in everyday language, identify the outcomes in the sample space which compose the event.
  c. Design and use a simulation to generate frequencies for compound events.

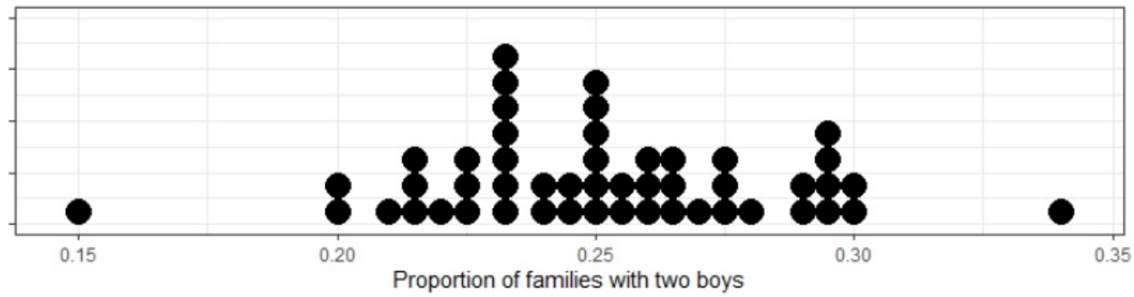
Let's assume that when a family has children that each child is equally likely to be a boy or girl. If a family has only one child, then we will denote the possible outcomes as B and G with both outcomes equally likely. If a family has two children, then we have four possible outcomes for the gender of the children in their birth orders, $\{BB, BG, GB, GG\}$. Each of these events are also equally likely. Another way to represent this information is through a tree diagram that lists all of the possible events and the probabilities of those events.



In a simulation of 200 families, we have the following probabilities:

BB: 0.26 BG: 0.20 GB: 0.26 GG: 0.28

When we look at a sample of 50 simulations of 200 families, we have the following distribution for the proportion of the families with two boys. The sample had a mean of 0.2511 and standard deviation of 0.0332.



As we increase the size of our simulation and/or the number of simulations, we will find that the probabilities for each of the events will get closer to the theoretical probabilities found using the tree diagram.

Related Content Standards

- (7.SP.7) Develop a probability model and use it to find probabilities of events. Compare probabilities from a model to observed frequencies; if the agreement is not good, explain possible sources of the discrepancy.
  a. Develop a uniform probability model by assigning equal probability to all outcomes, and use the model to determine probabilities of events.
  b. Develop a probability model (which may not be uniform) by observing frequencies in data generated from a chance process.

The tree diagram method of organizing the possible events and their probabilities is particularly useful when the probabilities are not all the same. In a deck of cards, there are 52 cards with 12 of them being face cards. Let's explore the probabilities related to the number of face cards drawn in when drawing three cards. In the following tree diagram we will denote the drawing of a face card by **F** and the drawing of a number card by **N**. Since we will be drawing cards and not replacing them, the probability of drawing face cards and number cards change as we draw the cards. For instance, if we draw a face card when we go to draw the next card there are a total of 51 cards to choose from and only 11 face cards left in the deck.

15.1.2 Estimating Probabilities

There are many times where we cannot determine the theoretical probability of an event, or doing so would be very difficult. In these cases, we frequently use a series of experiments or simulations to determine the long-run relative frequency of the event to approximate the probability of the event. In sports, we estimate a player's probability of making a free throw, kicking a field goal, or hitting a home run by looking at the long-run relative frequency of those events in their past performances, either during the current game or throughout the season.

Related Content Standards

- (7.SP.6) Approximate the probability of a chance event by collecting data on the chance process that produces it and observing its long-run relative frequency, and predict the approximate relative frequency given the probability.

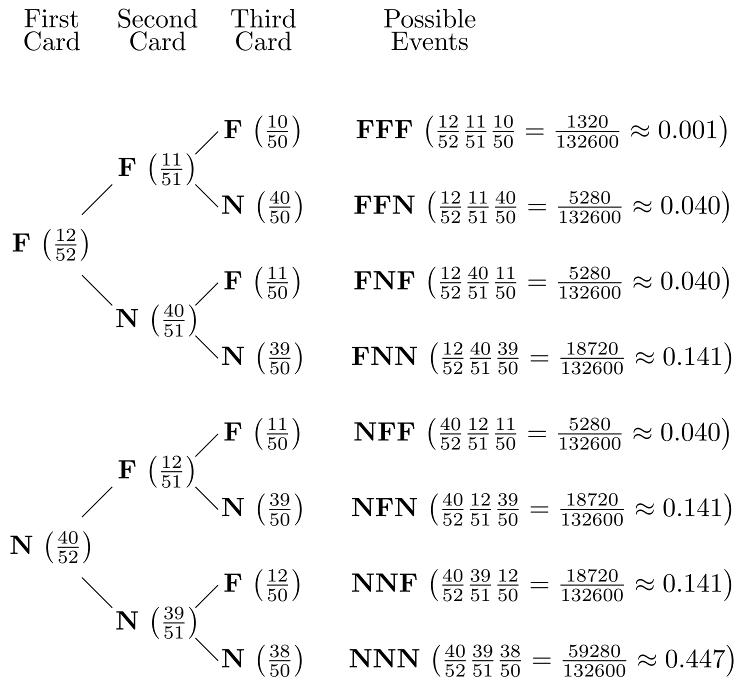


Table 15.2:

Number of hits in a game	0	1	2	3	4
Number of games with that many hits	150	206	113	30	1
Probability of that many hits	30%	41.2%	22.6%	6%	0.2%

Example 15.1. Assume that a baseball player typically has four at-bats in a baseball game. If his hitting percentage is .253, what is the probability of having at least three hits during the game?

We will see in a later section how to estimate this probability using theoretical probability techniques, but this can also be estimated using simulations.

Using the random number generator on a spreadsheet (`=randbetween(1,1000)`) one can designate a number less than or equal to 253 as a hit. We can do this four times to simulate a game of four at-bats. We can then create 500 of these possible games to estimate how many hits this player is likely to get.

The following "Birthday problem" was posed by Richard von Mises in 1939 (https://en.wikipedia.org/wiki/Birthday_problem):

> How many people would need to be in a room so that it was more likely than not that at least two of them shared a birthday?

We can assume that all days of the year are equally likely to be a birthday and estimate this using a series of simulations by using a random number generator to generate a number between 1 and 365 to randomly generate a birthday. We can then choose different numbers of people in a room to estimate the probability of two people in the room sharing a birthday. Using simulations[1] of 500 rooms we can estimate the probabilities of at least two people sharing a birthday for different numbers of people in the room.

[1] https://www.r-bloggers.com/the-birthday-paradox-puzzle-tidy-simulation-in-r/

Table 15.3:

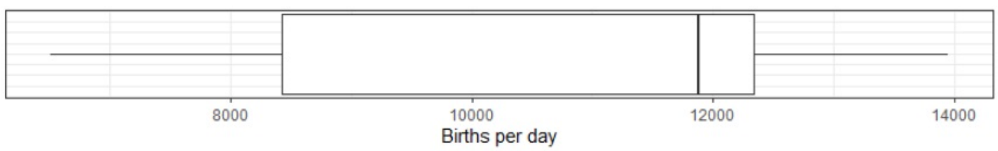
People in the room	10	20	21	22	23	24	25	30	40
Probability of a shared birthday	13.8%	37.4%	46.2%	49.8%	53%	53.8%	60.8%	70.4%	88%

Table 15.4:

People in the room	21	22	23	24	25
Probability of a shared birthday	45.7	49.0%	52.6%	54.9%	58.3%

We can see that we would need to have around 22 or 23 people in the room to have it more likely than not that two people share a birthday.

However, if we look at the number of people born in the United States in 2015[2] we see that there is a significant variability in the number of people born each day.



If we use the birth data from 2015 to create 10,000 simulations at each room population from 20 to 30 we see that it still is around 22 people in a room to likely have two people share a birthday.

There are also times where events are not repeatable, but involve uncertainty and it is still reasonable to estimate the probability of the event. These include the probability of a certain team winning the national championship this year or a certain candidate winning a presidential election. These are often referred to as **subjective probability**.

A common method to determine a subjective probability is the combination of observed data and statistical models to generate simulations that create a **probability forecast** of an event occurring. The probability of precipitation (or chance of rain) uses data collected from various weather stations to create models for the weather. These models are then analyzed with simulations to generate a probability that some minimum quantity of precipitation will occur within a specified forecast period and location.

The process of determining the probability of a team winning a national championship involves using predefined data, estimates of winning percentages for each game, and then running thousands of simulations of the entire season and reporting the percentage of those simulations that result in a championship.

15.1.3 Exercises

1. Gumballs[3] - Imagine a gumball machine with 10 gumballs that always has 4 blue, 3 yellow, 2 green, and 1 red gumball. So as you remove a gumball from the machine it magically replaces it with the color you removed so that it always has the same proportion of colors.

    a. Create a physical (with marbles or cards) or computer simulation to answer the question, "If you took 10 gumballs out of the machine, how many of each color gumballs would you get?"

[2]Using the MosaicData package in R
[3]Adapted from https://www.amstat.org/asa/files/pdfs/stew/TheGumballMachine.pdf

b. How would the answer change if you had 50 people take out 10 gumballs?
c. How would your answer change if 50 people took out 20 gumballs instead?

2. In the game of "Pass the Pigs" plastic pigs are rolled instead of dice. Since these are non-standard dice we do not know the probability of how each pig will land and so we need to run experiments to estimate the probability. Write up a plan for how you could have a class run experiments to estimate the probability for each of the possible ways for a pig to land.

15.2 Probability Spaces

Throughout this section we will define many different terms and develop a great deal of notation regarding the theory of probability. In order to assist us in this process we will use the game of Pig as a concrete example to study.

There are many different variations of the game of Pig, but the one that we will begin with involves the rolling of a single die. For each turn, a player repeatedly rolls a die until either a 1 is rolled or the player decides to 'hold'. If the 1 is rolled, the player scores nothing on that turn and it becomes the next player's turn. If a 1 is not rolled, the player chooses to either roll again or 'hold'. If the player chooses to 'hold', that player receives a score for that turn of the sum of the rolls up to that point. This score on the turn is then added to the player's previous score to create the total score. The players then take turns until one player has a total score of 100 points.

In order to better understand this process we will create an example game.

- *Player 1:* Rolls a 3 and decides to roll again and rolls a 1. This player gets no points for the turn.
- *Player 2:* Rolls a 4, then a 5, then a 2 and decides to hold. This player gets 11 points for the turn.
- *Player 1:* Rolls a 6 and a 2 and decides to hold with 8 points for the turn.
- *Player 2:* Rolls a 3 and a 2 and decides to hold. Player 2 now has 5 points for the turn and 16 total points.
- *Player 1:* Rolls a 1. Player 1 gets 0 points for the turn and has 8 total points.
- *Player 2:* ...

So for each roll, a player has six possible outcomes on the die, $\{1, 2, 3, 4, 5, 6\}$. This set of possible outcomes is called the sample space of the random phenomenon of rolling a die.

Definition 15.1. The **sample space**, Ω, is the set of all possible outcomes of a random phenomenon. An **outcome**, ω, is an element of the sample space ($\omega \in \Omega$).

An **event** A is a subset of the sample space, $A \subseteq \Omega$. If the random phenomenon give the outcome ω, we say that event A occurred if $\omega \in A$.

In the game of Pig we can consider the rolls in a turn to be the sample space under consideration. In this situation, each element of the sample space would be a sequence of rolls of the die, with all but a finite number being blank. So one element of this sample space would be

$$\omega = (2, 3, 6, 5, 1, ., ., \ldots)$$

if a person rolled a 2, then a 3, then a 6, then a 5, then a 1. After the roll of the 1, there would not be any more rolls.

Now that we have the idea of a sample space we are ready to define the probability for events in that sample space.

Definition 15.2. Let Ω be a sample space of outcomes and let $\mathcal{F}$ be a collection of subsets of Ω that satisfy the following (σ-field) properties:

- (S1) $\Omega \in \mathcal{F}$
- (S2) If $A \in \mathcal{F}$, then $A^c = \Omega \setminus A \in \mathcal{F}$.
- (S3) If I is a finite or countably infinite indexing set and $\{A_i\}_{i \in I} \subset \mathcal{F}$, then $\bigcup_{i \in I} A_i \in \mathcal{F}$.

A function $P : \mathcal{F} \to [0, 1]$ is called a **probability measure** if it satisfies the following properties:

- (P1) $P(\Omega) = 1$ and $P(\emptyset) = 0$
- (P2) For all events $A \in \mathcal{F}$, $0 \leq P(A) \leq 1$.
- (P3) If events $A_1, A_2, \ldots \in \mathcal{F}$ are mutually disjoint ($A_i \cap A_j = \emptyset$ for all $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

The σ-field properties of $\mathcal{F}$ ensure that the corresponding properties of P make sense. Most of the sample spaces encountered in the K-12 curriculum are finite spaces and the collection of subsets, $\mathcal{F}$, of Ω is usually just the power set of Ω.

If Ω is a non-empty set, we know that $\Omega \in \mathcal{F}$ and property (S2) says that $\emptyset \in \mathcal{F}$. So the smallest collection of subsets of Ω that satisfy these conditions is $\{\emptyset, \Omega\}$.

Using the generalized De Morgan's Laws (Theorem 2.7)

$$\bigcap_{i \in I} A_i = \left(\bigcup A_i^c\right)^c$$

we can combine properties (S2) and (S3) to see that the intersections of sets in $\mathcal{F}$ is also in $\mathcal{F}$.

The requirement that $P(\Omega) = 1$ ensures that the probability of all possible outcomes is 1. Similarly, $P(\emptyset) = 0$ is that the option of no outcome occurring is not possible.

The requirement that

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

when the A_i are mutually disjoint is often referred to as the countable additivity of the probability measure. Another way of describing mutually disjoint events is mutually exclusive.

Definition 15.3. Two events are **mutually exclusive events** if they cannot happen at the same time.

$$A \cap B = \emptyset$$

In our example of playing Pig with only rolling two times on each turn we can rewrite some of our information using this new terminology. We can consider Ω to be the possible pairs of rolls,

$$\Omega = \{(a, b) | a, b \in \{1, 2, 3, 4, 5, 6\}\}$$

and $\mathcal{F} = \mathcal{P}(\Omega)$. Then the probability of rolling a 1 for at least one of the rolls would be

$$P(\{(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (3,1), (4,1), (5,1), (6,1)\}) = \frac{11}{36}$$

and so the probability of getting zero points for a turn using the two rolls strategy is $\frac{11}{36}$.

15.2.1 Boolean Algebra and Probability

Since P maps subsets of Ω into the interval $[0,1]$ we need to understand how P operates with the Boolean algebra of set theory and how that relates to the language of probability.

We first note that for two sets $A, B \subseteq \Omega$, $A \cap B$ is the set of outcomes in Ω that are in both A and B. In our example from the game of Pig with rolling twice for a turn, we can let A be the set of events where a 3 is rolled on the first roll. So

$$A = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)\} \text{ and } P(A) = \frac{6}{36}.$$

Let B be the set of events where a 4 or 5 is rolled on the second roll,

$$B = \{(1,4),(2,4),(3,4),(4,4),(5,4),(6,4),(1,5),(2,5),(3,5),(4,5),(5,5),(6,5)\} \text{ and } P(B) = \frac{12}{36}.$$

Then $A \cap B$ is the set of events where a 3 is rolled on the first roll and a 4 or 5 is rolled on the second roll,

$$A \cap B = \{(3,4),(3,5)\} \text{ and } P(A \cap B) = \frac{2}{36}.$$

Related Content Standards

- (HSS.CP.1) Describe events as subsets of a sample space (the set of outcomes) using characteristics (or categories) of the outcomes, or as unions, intersections, or complements of other events (“or,” “and,” “not”).

Similarly, $A \cup B$ is the set of events that occur in either A or B, or both. In our example above,

$$A \cup B = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(1,4),(2,4),(4,4),(5,4),(6,4),(1,5),(2,5),(4,5),(5,5),(6,5)\}$$

and $P(A \cup B) = \frac{16}{36}$.

We can notice that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

The following theorem generalizes this result and describes how probability measures interact with Boolean algebra of sets.

Theorem 15.1. *If $A, B \in \mathcal{F}$ and $P : \mathcal{F} \to [0,1]$ is a probability measure, then*

a. $P(A^c) = 1 - P(A)$
b. $P(A \setminus B) = P(A) - P(A \cap B)$
c. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
d. If $A \subseteq B$, then $P(A) \leq P(B)$.

Proof. **Proof of (a):** Since $A \in \mathcal{F}$, property (S2) says that A^c is also in $\mathcal{F}$ and so $P(A^c)$ is well-defined. We also know that $\Omega = A \cup A^c$ and that $A \cap A^c = \emptyset$. So property (P3) states that

$$P(\Omega) = P(A) + P(A^c).$$

Since $P(\Omega) = 1$, we have that $P(A^c) = 1 - P(A)$.

Proof of (b): Let $A, B \in \mathcal{F}$. Since $\Omega = B \cup B^c$, we know that

$$A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c) = (A \cap B) \cup (A \setminus B).$$

Since $(A \cap B)$ and $A \setminus B$ are disjoint, property (P3) gives us that $P(A) = P(A \cap B) + P(A \setminus B)$. We can rearrange this equation so that

$$P(A \cap B) = P(A) - P(A \setminus B).$$

Proof of (c): Let $A, B \in \mathcal{F}$. Then we know that

$$(A \setminus B) \cup (A \cap B) \cup (B \setminus A)$$

is a partition of $A \cup B$ in that the three sets are mutually disjoint and their union is $A \cup B$. So

$$P(A \cup B) = P(A \setminus B) + P(A \cap B) + P(B \setminus A)$$

by property (P3). Combining this with part (b) of the theorem we have

$$P(A \cup B) = (P(A) - P(A \cap B)) + P(A \cap B) + (P(B) - P(A \cap B)) = P(A) + P(B) - P(A \cap B).$$

Proof of (d): Let $A, B \in \mathcal{F}$ such that $A \subseteq B$. Since $A \subset B$,

$$B = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^c) = A \cup (B \setminus A)$$

and A and $B \setminus A$ are disjoint. So property (P3) gives us that $P(B) = P(A) + P(B \setminus A)$. Since $P(B \setminus A) \geq 0$, $P(A) \leq P(B)$. $\square$

Related Content Standards

- (HSS.CP.7) Apply the Addition Rule, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, and interpret the answer in terms of the model.

A common example used to better understand the usefulness of the language of probability is called the 'Monty Hall' problem as it is based on a scenario from the game show 'Let's Make a Deal' hosted by Monty Hall. In this scenario there are three doors in the game show studio. Behind one of the doors is a new car. Behind the other two doors is a goat. The contestant is asked to choose which door they think the car is behind. Monty then reveals the goat behind one of the doors that was not chosen and asks the contestant if they would like to stick with their current door or change to the new door. Because of the condition and event that one of the doors has been revealed, we see that the probability of a car behind each door has changed.

We will denote the sample space in this problem as

$$\Omega = \{CGG, GCG, GGC\}$$

where CGG denotes that the car is behind Door 1. Assume that a contestant originally chooses Door 2. Since $P(\{GCG\}) = \frac{1}{3}$ their probability of winning the car is $\frac{1}{3}$. Monty then reveals that the car is not behind Door 3 so that $P(\{GGC\}) = 0$. This means that $P(\{CCG\}) = \frac{2}{3}$, meaning that the player should definitely change their choice of doors because it doubles their chance of winning a car.

We can verify this conclusion by assuming that the contestant chooses Door 1 and look at the possible outcomes.

15.2.2 Exercises

1. How likely is it that a family with five children has all boys or all girls?

   a. Answer the question by listing out all of the possible combinations of boys and girls.
   b. Answer the question using a simulation and estimating the probability.
   c. Answer the question using properties of compound events.
   d. Compare and contrast the different methods.
   e. How would the different methods perform with changing the question to 'How likely is it that a family with five children has exactly two girls?' or 'How likely is it that a family with five children has at least two girls?'

Table 15.5:

Behind Door 1	Behind Door 2	Behind Door 3	Result if Staying	Result if Switching
Car	Goat	Goat	Wins Car	Loses
Goat	Car	Goat	Loses	Wins Car
Goat	Goat	Car	Loses	Wins Car

15.3 Conditional Probability

There are times that we want to know the probability of a set of events conditioned on the occurrence of another set of events. For instance, if we are rolling two dice, we may want to know the probability that the sum of the dice is greater than 4 conditioned on the first die being a 2. With this scenario we have reduced the overall sample space to just those events for which the first die is a two. So what we are wanting to know is

$$\frac{P(\text{sum of dice is greater than 4})}{P(\text{first die is a two})} = \frac{4}{6}$$

since the total possible events are $\{(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)\}$ with only 4 of the 6 being the desired events.

Definition 15.4. If $A, B \in \mathcal{F}$ then the **conditional probability of** A **given** B is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ when } P(B) > 0.$$

Similarly,

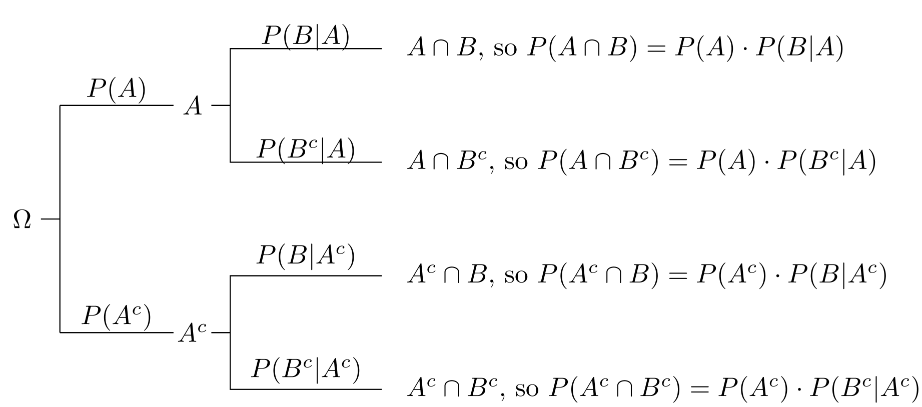
$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ when } P(A) > 0.$$

The assumption that $P(B) > 0$ for $P(A|B)$ is reasonable in that if $P(B) = 0$, we would be assuming that an impossible event occurs.

Theorem 15.2. *If $A, B \in \mathcal{F}$, with $P(A) > 0$ and $P(B) > 0$, then*

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

This theorem is demonstrated in the following tree.



Definition 15.5. Two events are **independent** if the occurrence of one does not affect the probability of occurrence of the other.

If two events are not independent, they are called **dependent**.

In terms of the probability measures, events A and B are independent if and only if

$$P(A|B) = P(A) \quad \text{and} \quad P(B|A) = P(B)$$

since the probability of event A should be the same whether or not event B has occurred, similarly with $B|A$. Combining this with the previous theorem we have the following theorem.

Theorem 15.3. *Two events A and B are independent if and only if their joint probability equals the product of their probabilities.*

$$P(A \cap B) = P(A)P(B)$$

Proof. Assume that A and B are independent events. So $P(A|B) = P(A)$. Combining this with Theorem 15.2 we have that

$$P(A \cap B) = P(B) \cdot P(A|B) = P(B) \cdot P(A).$$

If we instead assume that $P(A \cap B) = P(A) \cdot P(B)$, Theorem 15.2 implies that

$$P(A)P(B) = P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

and so we have

$$P(B|A) = P(B) \text{ and } P(A|B) = P(A).$$

So the events are independent. □

Related Content Standards

- (HSS.CP.2) Understand that two events A and B are independent if the probability of A and B occurring together is the product of their probabilities, and use this characterization to determine if they are independent.
- (HSS.CP.3) Understand the conditional probability of A given B as P(A and B)/P(B), and interpret independence of A and B as saying that the conditional probability of A given B is the same as the probability of A, and the conditional probability of B given A is the same as the probability of B.

Let's return to our example of drawing three cards and determining the probabilities of draw a face card or a number card.

Let A be the events where the first card drawn was a face card and B be the events where the third card drawn was a face card. So

$$A = \{FFF, FFN, FNF, FNN\} \text{ and } B = \{FFF, FNF, NFF, NNF\}.$$

Since each of the events in A are mutually exclusive we have that

$$\begin{aligned} P(A) &= P(\{FFF\}) + P(\{FFN\}) + P(\{FNF\}) + P(\{FNN\}) && (15.1)\\ &= \frac{1320 + 5280 + 5280 + 18720}{132600} && (15.2)\\ &= \frac{17}{221} = \frac{3}{13}, && (15.3) \end{aligned}$$

First Card	Second Card	Third Card	Possible Events
F $\left(\frac{12}{52}\right)$	**F** $\left(\frac{11}{51}\right)$	**F** $\left(\frac{10}{50}\right)$	**FFF** $\left(\frac{12}{52}\frac{11}{51}\frac{10}{50} = \frac{1320}{132600} \approx 0.001\right)$
		N $\left(\frac{40}{50}\right)$	**FFN** $\left(\frac{12}{52}\frac{11}{51}\frac{40}{50} = \frac{5280}{132600} \approx 0.040\right)$
	N $\left(\frac{40}{51}\right)$	**F** $\left(\frac{11}{50}\right)$	**FNF** $\left(\frac{12}{52}\frac{40}{51}\frac{11}{50} = \frac{5280}{132600} \approx 0.040\right)$
		N $\left(\frac{39}{50}\right)$	**FNN** $\left(\frac{12}{52}\frac{40}{51}\frac{39}{50} = \frac{18720}{132600} \approx 0.141\right)$
N $\left(\frac{40}{52}\right)$	**F** $\left(\frac{12}{51}\right)$	**F** $\left(\frac{11}{50}\right)$	**NFF** $\left(\frac{40}{52}\frac{12}{51}\frac{11}{50} = \frac{5280}{132600} \approx 0.040\right)$
		N $\left(\frac{39}{50}\right)$	**NFN** $\left(\frac{40}{52}\frac{12}{51}\frac{39}{50} = \frac{18720}{132600} \approx 0.141\right)$
	N $\left(\frac{39}{51}\right)$	**F** $\left(\frac{12}{50}\right)$	**NNF** $\left(\frac{40}{52}\frac{39}{51}\frac{12}{50} = \frac{18720}{132600} \approx 0.141\right)$
		N $\left(\frac{38}{50}\right)$	**NNN** $\left(\frac{40}{52}\frac{39}{51}\frac{38}{50} = \frac{59280}{132600} \approx 0.447\right)$

$$P(B) = P(\{FFF\}) + P(\{FNF\}) + P(\{NFF\}) + P(\{NNF\}) \tag{15.4}$$
$$= \frac{1320 + 5280 + 5280 + 18720}{132600} \tag{15.5}$$
$$= \frac{17}{221} = \frac{3}{13}, \tag{15.6}$$

and

$$P(A \cap B) = P(\{FFF, FNF\}) = \frac{1320 + 5280}{132600} = \frac{11}{221}.$$

From these we have that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{11}{17} \neq P(A)$$

and

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{11}{17} \neq P(B)$$

and so these are not independent events.

The non-independence of these events makes sense in this situation since the probability that the third card is a face card is dependent upon the first card being a face card as that would reduce the number of face cards that could be drawn later.

Because of the vagueness of the English language, individuals often are confused about the differences between independent and mutually exclusive events. Recall that two events, A and B, are mutually exclusive if they cannot both occur at the same time, $A \cap B = \emptyset$. This means that two events are mutually exclusive if $P(A|B) = 0$ and $P(B|A) = 0$.

Related Content Standards

Table 15.6:

	Not Proficient	Proficient	Totals
Male	425	299	724
Female	362	295	657
Totals	787	594	1381

Table 15.7:

	Not Proficient	Proficient	Totals
Male	0.308	0.217	0.524
Female	0.262	0.214	0.476
Totals	0.570	0.430	1.000

- (HSS.CP.4) Construct and interpret two-way frequency tables of data when two categories are associated with each object being classified. Use the two-way table as a sample space to decide if events are independent and to approximate conditional probabilities.
- (HSS.CP.5) Recognize and explain the concepts of conditional probability and independence in everyday language and everyday situations.

We will now look at the math test scores for eighth graders in a local school district to determine if proficiency in math, as measured by the test, is dependent or independent of Gender and/or Race using some two-way frequency tables.

From this table we see that $P(\text{Male}) = \frac{724}{1381} \approx 0.524$, $P(\text{Proficient}) = \frac{594}{1381} \approx 0.430$, and

$$P(\text{Male and Proficient}) = \frac{299}{1381} \approx 0.217.$$

Since

$$P(\text{Male}) \cdot P(\text{Proficient}) \approx 0.524 \cdot 0.430 \approx 0.225$$

we see that it is very likely that proficiency on this math test is independent of gender as 0.225 is fairly close to 0.217.

These results can also be seen in the joint relative frequency table where independence is checked to see if the values inside the table match the product of the marginal probabilities.

We now turn to a comparison of White and Black students in the same eighth grade math assessment with a two-way frequency table.

We now look at the joint relative frequency table to determine independence.

Table 15.8:

	Not Proficient	Proficient	Totals
Black or African American	324	107	431
White	388	465	853
Totals	712	572	1284

Table 15.9:

	Not Proficient	Proficient	Totals
Black or African American	0.252	0.083	0.336
White	0.302	0.362	0.664
Totals	0.555	0.445	1.000

Table 15.10:

	Not Proficient	Proficient	Totals
Black or African American	0.752	0.248	1
White	0.455	0.545	1

Since $P(\text{White}) \cdot P(\text{Proficient}) = 0.664 \cdot 0.445 = 0.296$ and $P(\text{White and Proficient}) = 0.362$ we see that proficiency is not independent of race. Instead, we find

$$P(\text{Proficient}|\text{White}) = \frac{P(\text{Proficient and White})}{P(\text{White})} \approx \frac{0.362}{0.664} \approx 0.545$$

and

$$P(\text{Proficient}|\text{Black}) = \frac{P(\text{Proficient and Black})}{P(\text{Black})} \approx \frac{0.083}{0.336} \approx 0.247$$

which can be seen in the relative frequency table.

Related Content Standards

- (HSS.CP.6) Find the conditional probability of A given B as the fraction of B's outcomes that also belong to A, and interpret the answer in terms of the model.
- (HSS.CP.8) Apply the general Multiplication Rule in a uniform probability model, $P(A \text{ and } B) = P(A)P(B|A) = P(B)P(A|B)$, and interpret the answer in terms of the model.

15.3.1 Exercises

1. Use a simulation of 1000 pairs of coin flips. Use this data set to determine if Heads on the second flip is independent of Heads on the first flip.
2. Find a data set of interest (similar to the education example) and determine if two variables are dependent or independent.

15.4 Random Variables and Expected Value

Definition 15.6. Let $(\Omega, \mathcal{F}, P)$ be a probability space.A **random variable (RV)** on the probability space is a real-valued function $X : \Omega \to \mathbb{R}$.

We can now compute the probability of certain values occurring for the random variable by defining,

$$P(X = \alpha) := P(\{\omega \in \Omega | X(\omega) = \alpha\})$$

Table 15.11:

ω	1	2	3	4	5	6
‘roll_points‘	0	P_0+2	P_0+3	P_0+4	P_0+5	P_0+6

and

$$P(X \leq \beta) := P(\{\omega \in \Omega | X(\omega) \leq \beta\})$$

and so on.

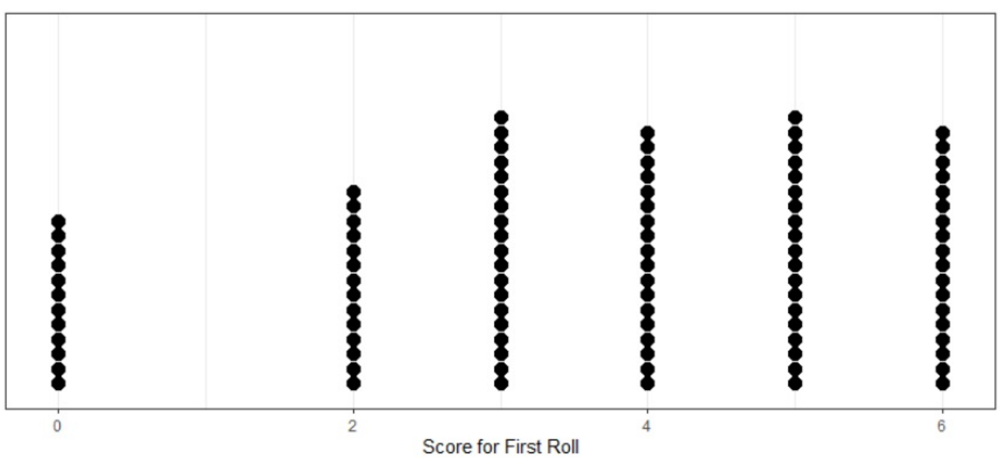
If we consider the sample space, Ω_1, of rolling a die during a turn in the game of Pig have have that

$$\Omega_1 = \{1, 2, 3, 4, 5, 6\}$$

and we can define a function, $X_1 : \Omega_1 \to \mathbb{R}$ using the following table, where P_0 is the amount of points in that turn prior to that roll.

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 2 > 1' in coercion to 'logical(1)'
```

In particular, for the first roll, the possible outcomes of 0,2,3,4,5, or 6 points are equally likely. So the expected score after the first roll is the average of these numbers, $\frac{20}{6}$, even though there is actually no way to get that score. If we run a simulation with 100 rolls for this situation, we see that each of the outcomes have approximately the same probability of occurring.



If we are in the situation where we have already rolled the die and have points, if we choose to roll during a turn we could expect to earn the average of the possible points on the roll, since all of the possibilities have equal probability. So the expectation is that by rolling again we would have the score of

$$\frac{0 + (P_0 + 2) + (P_0 + 3) + (P_0 + 4) + (P_0 + 5) + (P_0 + 6)}{6} = \frac{5P_0 + 20}{6}$$

Related Content Standards

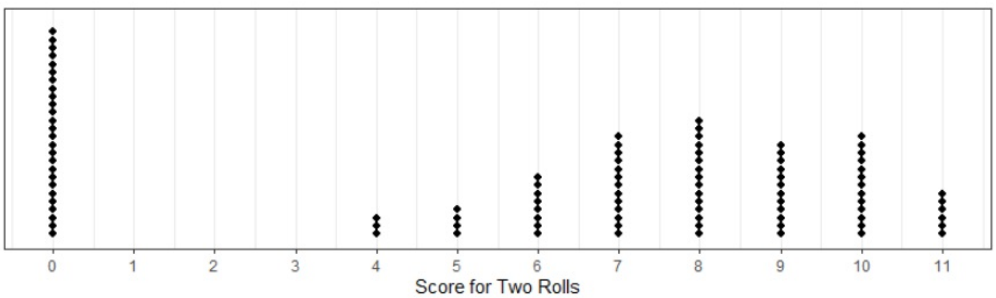
- (HSS.MD.2) Calculate the expected value of a random variable; interpret it as the mean of the probability distribution.

Table 15.12:

Probability value of ‘turn_points‘	0	4	5	6	7
Probability of that score in A	$\frac{12}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36

If we consider the sample space, Ω_t, of a player's turn in the game of Pig the number of elements in the sample space is much larger. We can also define a function $X_t : \Omega_t \to \mathbb{R}$ where $X_t(\omega)$ is the amount of points corresponding with that sequence of dice rolls for a turn. We will say that `turn_points` $= X_t(\omega)$.

Suppose that we let $A \subset \Omega_t$ be the possible rolls if we limit a turn to a maximum of two rolls. We have the following possible outcomes for `turn_points` for $\omega \in A$ with a dot plot of 100 simulations that further shows the distribution of possible scores.



Related Content Standards

- (HSS.MD.1) Define a random variable for a quantity of interest by assigning a numerical value to each event in a sample space; graph the corresponding probability distribution using the same graphical displays as for data distributions.

From this sample of 100 simulations, we see that the average value of the random variable is 5.95. Using a sample of 1000 simulations, we find that the average value of the random variable of the score after two rolls is 5.421. If we look at all of the possible pairs of rolls for two dice we can determine the probability of getting each possible score value. Since there are 36 different possible rolls for the two dice, we have the following probability distribution.

We can then determine the expected value for `turn_points` with the assumption that there are two rolls by taking a weighted mean.

$$\text{Expected Value of a Random Variable} = \sum (\text{possible value}) \cdot (\text{probability of that value})$$

$$\begin{aligned} E(X) &= 0 \cdot \frac{12}{36} + 4 \cdot \frac{1}{36} + 5 \cdot \frac{2}{36} + 6 \cdot \frac{3}{36} + 7 \cdot \frac{4}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} && (15.7) \\ &= \frac{4 + 10 + 18 + 28 + 40 + 36 + 30 + 22 + 12}{36} = \frac{200}{36} && (15.8) \\ &= 5 + \frac{5}{9} && (15.9) \end{aligned}$$

Related Content Standards

- (HSS.MD.3) Develop a probability distribution for a random variable defined for a sample space in which theoretical probabilities can be calculated; find the expected value.

We can also use this terminology to rewrite the expression for the expected value of a discrete random variable X,

$$E(X) = \sum_x x \cdot P(X = x).$$

If X is a continuous random variable with a probability density function, $f(x)$, the summation becomes an integral and

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx.$$

Related Content Standards

- (HSS.MD.5) Weigh the possible outcomes of a decision by assigning probabilities to payoff values and finding expected values.

  a. Find the expected payoff for a game of chance.
  b. Evaluate and compare strategies on the basis of expected values.

Returning to our game of Pig, recall that the expected value of the score on a turn when rolling the die one time is 3.667 and is 5.556 when rolling two times. So it seems that the expected value of the score increases as we increase the number of times that we roll the dice on each turn. However, if we think about rolling the dice 20 times on a turn we are almost certain that at least one of those 20 rolls will include a 1 making the expected turn score to be close to 0. This implies that there must be some number of rolls on each turn that would maximize the expected score for a turn. Because the theoretical probabilities become more complicated as we increase the number of rolls on a turn, we can obtain a good estimate for the expected score using 1000 simulations for each number of turns.

From these simulations we can see that the best strategy, in terms of the number of rolls each turn, is to roll around 5 or 6 times each turn. We could improve our estimates for the expected value by increasing the number of simulations.

Another option is to run the simulations many times to get a better understanding of the expected values for a turn when rolling 4, 5, 6, 7, or 8 times for a turn. We report the results of 100 experiments of 1000 simulations for each of the options. The mean of the scores for the experiments is equivalent to a mean of a simulation of 100,000 trials.

However, we can see from the distributions of these experiments in the box plots below that the expected scores of the 1,000 simulations had some variability.

This means that we cannot definitively say if five or six rolls per turn is the best strategy in terms of the number of rolls per turn.

15.4.1 Exercises

1. Find the best strategy for Pig based on stopping your turn once you have reached a certain number of points for that turn.

2. Use the theoretical probabilities, Boolean algebra, and conditional probability to determine the expected values for five and six rolls per turn.

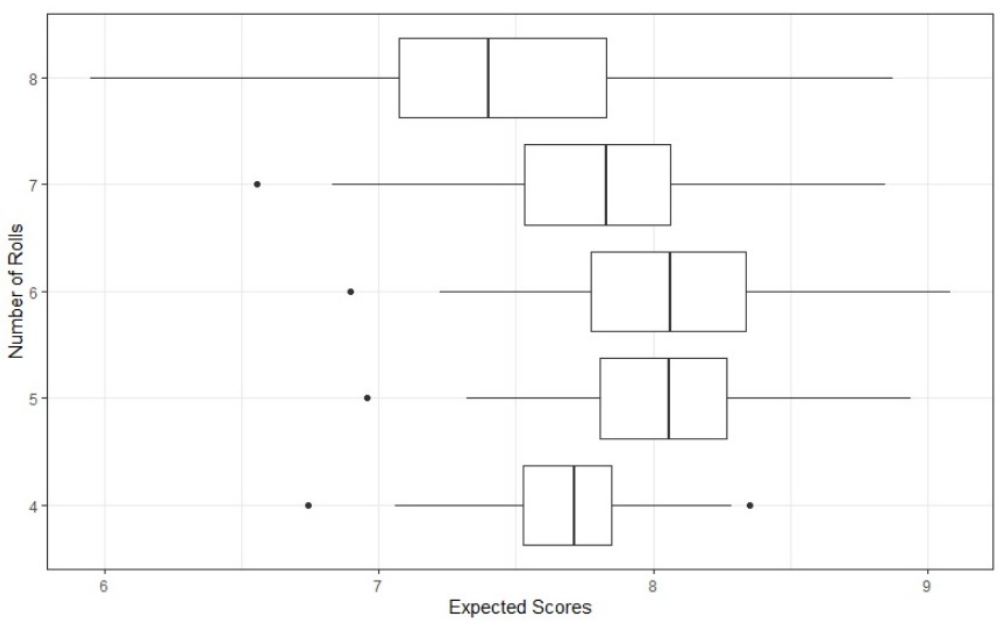
3. In the game of Risk, two players have battles by rolling different numbers of dice. For a battle, the attacker can roll up to 3 dice and the defender can roll up to two dice. Use simulations to analyze the expected value of the number of armies won or lost by the attacker and the defender based on the number of dice chosen by each person for the battle.

Table 15.13:

Rolls per Turn	Expected Value of Points per Turn (using 1000 simulations)
0	0.000
1	3.667
2	5.421
3	7.119
4	7.183
5	8.060
6	7.849
7	7.134
8	7.283
9	6.801
10	5.991
11	6.157
12	5.258
13	4.414
14	4.327

Table 15.14:

Rolls per Turn	Four	Five	Six	Seven	Eight
Mean Score	7.6978	8.0453	8.0562	7.8233	7.4375



Chapter 16

Estimating Parameters and Testing Hypotheses

There are many times where we would like to know something about a large population (height of a group of people, weight of a manufactured item, or income levels) but the collection of the information from the entire population is too difficult, time consuming, risky, or expensive. In these cases we would like to estimate the properties of the variable for the entire population by studying a small sample of the population. In this chapter we will learn different techniques for selecting a sample, analyzing the variables of the sample, and drawing inferences from the information to the entire population.

Related Content Standards

- (7.SPA.1) Understand that statistics can be used to gain information about a population by examining a sample of the population; generalizations about a population from a sample are valid only if the sample is representative of that population. Understand that random sampling tends to produce representative samples and support valid inferences.
- (7.SPA.2) Use data from a random sample to draw inferences about a population with an unknown characteristic of interest. Generate multiple samples (or simulated samples) of the same size to gauge the variation in estimates or predictions.
- (HSS.IC.1) Understand statistics as a process for making inferences about population parameters based on a random sample from that population.

16.1 Sampling Techniques and Study Types

16.1.1 Sampling Errors and Bias

Statistics would be theoretical mathematics if we did not have to operate in the real world. However, because statistics evolved out of a desire to make conclusions about the real world and the real world relies on humans, there are a variety of errors we introduce into the theoretical models that we have to account for when applying statistics to the questions we have posed.

- *Selection bias.* One of the biggest causes of error when applying statistics is a bias caused by the method of selecting the sample for the analysis. There are many types of sampling techniques that researchers use to simplify the data collection process, with some of the most common listed below.
    - Judgement (purposive) sampling or purposive samples

  - Snowball sampling
  - Quota sampling
  - Convenience (accidental, opportunity) sampling
  - Voluntary sampling
  - Consecutive Samples: Keep accepting responses until the desired number of samples is achieved.

- *Random sampling error.* Another large cause of error in statistics arises as part of the random sampling process. Even if everything is done perfectly, any time one chooses a sample to represent a larger population, there is error introduced. This type of error can never be removed, but can be controlled and understood by paying attention to the sample size and the sample's representation of the overall population represented.

- *Coverage error.* If a researcher wishes to study the opinions and preferences of registered voters they may choose to use a data base of phone numbers. Since not all voters have a listed phone number and particular demographics are more likely to block unknown numbers on their phones, and under-coverage error may occur. If the researcher instead uses a list of email addresses it is very likely that a person may be contacted multiple times for the same survey if they have multiple email addresses. This would result in a possible over-coverage error.

- *Measurement error.* Any time that we try to measure something (the width of an object, a person's opinion, a student's knowledge) we cannot be entirely precise do to problems with our measurement techniques and instruments. Since we cannot eliminate the errors causes by the issues, we try to understand and control for these errors. For instance, if someone is trying to determine a person's opinion about something, the way that the question is phrased can have a very significant effect on how the person answers. In order to minimize this effect pilot groups are used to check how different people interpret the questions and questions are then rephrased.

- *Processing error.* Sometimes errors arise in the transfer and handling of data (i.e. copying information between a form and a spreadsheet) in the data collection and analysis process. These errors can be reduced by creating detailed protocols and having multiple people checking the work. However, even with the best protocols and backup contingencies, some processing errors make it through.

- *Non-response or participation bias.* Since most people do not like taking polls or surveys, there is likely to be a non-response bias. It is important in the data analysis process to try to understand this bias and take it into account in the design and interpretation of the analysis.

The good news is that careful planning will help us minimize some of these. The bad news is that because we are humans working with questions that don't have known answers, we can never be certain whether we've addressed all the important sources of error. The job of someone who is setting out to answer questions using statistics is not to eliminate all sources of errors, but rather to acknowledge those sources, explain the procedures they have done to minimize the potential impact they may have on the results, and discuss their findings in light of these potential error sources.

16.1.2 Probability Sampling

As our previous discussion shows, there are a lot of ways that we can bias a sample unintentionally by not doing a good job of picking our sample to reflect our population of interest. Statisticians have a solution to this that you have probably heard of called random sampling. Randomness (or probability sampling) introduces a way to gain a sample that is likely to be reflective of your population, but this is not as simple as it sounds–there are different ways to sample randomly, each which help address different bias' you may want to avoid in your sample.

To understand these, imagine you want to understand the citizens opinions about environmental regulation in a large city. The city has N people (your population) and you plan to sample n of them (where $n \leq N$).

Simple random sampling occurs when each person in the population (here, the city's citizens) has the same probability of being in your sample and this probability is $\frac{n}{N}$. With this method, we will have a sample that reflects our population.

Systematic sampling is similar, but instead of selecting your n people at random from the population, you apply some order to the population and then pick every kth person, where k is chosen to ensure the correct sample size at the end of the list. So, for our question, an order may be placed by listing all residential addresses, with a survey team member knocking on every 100th address in order to get a sample of 10 percent of the city's population.

Both simple random sampling and system sampling provide a random sample, but assume the population is uniform across all the ways that might be important to your question. If there are reasons to expect that is not the case, then, depending on how big n is relative to N, these techniques may not include people who are members of small groups that are important to you, but not likely to carry a lot of weight in a small sample.

For example, say your city has 100,000 people, and you sample 200 of them. If you are particularly interested in how families with children under 5 years old feel about your question, but only about 5 percent of your city's population are in this situation, you can expect that about 10 of your 200 people sample meet this criteria. This may not be enough representation of this particular group to have sufficient variance to understand their feelings about the environment.

You could resolve this issue by increasing your sample size, but collecting data tends to be costly, especially if you want a representative sample, so other ways have been developed to preserve randomness while increasing the representation of important categories within the sample.

Stratified random and **cluster sampling** are sampling techniques that follow this.

So, while "random" is good for sampling, it is not enough–you need to attend to the characteristics of how people are being chosen.

16.1.3 Study Types

Sampling, discussed in the previous section, does not occur in a vacuum: we collect data from a sample because we are conducting a study. The way that we collect data impacts the type of study we end up doing and the types of conclusions we can draw.

The most common type of study people learn about are experiments. Experiments are a type of study where we are able to randomly assign our sample into different conditions and then examine differences between the different conditions. There is a reason these are discussed widely–when we are able to control the environment of our sample, we have a pretty clear causal pathway and can thus make strong claims. Drug trials are often conducted using experiments for this reason. The FDA requires that drug developers show evidence of clear, positive effects when deciding whether a new drug is safe and effective or not. What characterizes an experiment is not randomness–most studies have an element of randomness built into them–but where the randomization occurs. In experiments, the randomness occurs when an individual is assigned into a condition. We may also pick participants randomly, but what is important is that once we have them, random assignment is used to determine which treatment they receive.

Related Content Standards

- (HSS.IC.3) Recognize the purposes of and differences among sample surveys, experiments, and observational studies; explain how randomization relates to each.

Studies without random assignment still can teach us a lot, but the causal pathway may be less clear. Many studies involve randomness, but randomness serves a different function within the study. For example, say we are interested in the relationship between the health of a population and the level of air pollution in their neighborhood. To investigate this question, we pick a random sample of people from a variety of neighborhoods and measure their health and the level of pollution in their neighborhood. This situation includes the selection of a random sample, but it is not an experiment because we do not have control over their conditions–we cannot, once a person is selected for our study, assign to them the condition of whether they have high or low pollution in their neighborhood–instead, we observe the variables of interest. Studies characterized by observing variables of interested on a sample are called observational studies. Note that observational studies are not defined by the presence of randomization, but , we often use sampling techniques such as stratified random sampling, to ensure we have enough variation in our sample to

A last common type of study is one in which a person uses a sample of convenience. For example, say an engineer is investigating the quality of widgets coming off the line of one of their machines. They use the 20 that came off the line first. This is a sample survey, but there is not randomization going on. These survey samples are common and characterized by the lack of randomization, either in terms of how the sample is collected or in how conditions are assigned. Such studies types are still valuable, but more in providing evidence that a pattern may exist. A more careful study that includes intentional sampling to better account for possible sources of variance and/or random sampling or assignment can improve and strengthen claims.

Note that the above discussion is about the role of randomness in terms of a study type. The sampling type also contributes to the characterization of a study and together, these elements build the study design.

16.1.4 Exercises

1. Write a summary of the various non-probabilistic sampling techniques listed at the beginning of this section.

16.2 Point and Interval Estimators

Related Content Standards

- (HSS.IC.4) Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

Suppose that we want to use the Census at School database to study the properties of secondary school students in the United States. While the sampling of the Census at School website is a sample of convenience, it is unlikely to be biased in the variable of height, but may be biased for other variables as this sample population is more likely to come from schools with higher socioeconomic levels.

Using a sample size of 1000 twelve-year-old students we find that the sample mean of the heights is 155.485 cm. This is used to estimate the population mean of the heights of twelve-year-old students in the United States. If we had used a different random sample of twelve-year-old students it is very possible that we would have had a different sample mean. So we would like to get an idea of how well 155.485 cm estimates the actual population mean that would be impossible to compute.

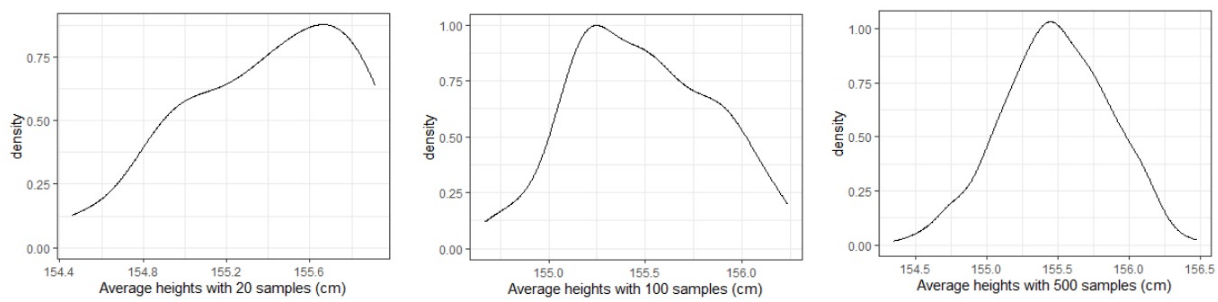
One way to measure how close the sample mean approximates the population mean is to use a simulation. We can randomly select 1000 data points from our sample data using replacements and take the mean of this data set. We can then do this 1000 times to create a distribution of possible means of the heights of twelve-year-old students. We include a sample of code for doing such a simulation in R.

```
nsims <- 1000 # number of simulations
m <- rep(NA,nsims) # empty vector to store means
for (i in 1:nsims){ m[i] <- mean(sample(Student_twelve$Height_cm, replace = TRUE))}
```

The result of this loop is a vector with 1000 means of the simulated sample heights.

We find that the mean of the simulations is also 155.47 cm and has a standard deviation of 0.364 0.365 cm. We also have the 95% of the simulations fell between 154.759 and 156.192. So we call the interval (154.759,156.192) a 95% confidence interval (95% CI) for the mean of the heights of the population. while this does not assure that the actual population mean is inside of this interval, we can be fairly certain that our interval contains the population mean.

We can notice that the density curve of the average heights of 1000 samples appears to be a normal distribution. If we look at similar density curves for average heights of 20, 100, and 500 samples we see that the shape of the curve is seeming to approach a normal distribution.



This property that the distribution of the averages of n samples approaches a normal distribution is the Central Limit Theorem.

Theorem 16.1 (Central Limit Theorem)**.** *If S_n is the sum of n mutually independent random variables, then the distribution function of S_n will approach a normal distribution as $n \to \infty$.*

So if our random variable of interest for the entire population, X, has a mean μ and standard deviation σ, the Central Limit Theorem implies that

$$\mu = E(\overline{X}) \quad \text{and} \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}.$$

16.2.1 Estimating Population Means and Medians

Using the Central Limit Theorem we can compute the standard error for the sample mean of the heights,

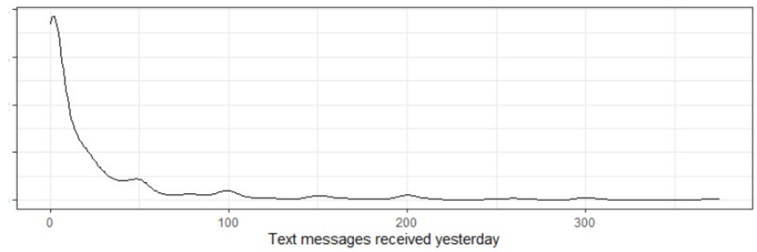
$$\sigma_{\overline{Y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{11.52035}{\sqrt{1000}} \approx 0.364\text{cm}.$$

We can also create a 95% confidence interval using the knowledge the 95% of the data in a normal distribution falls between 1.96 standard deviations below the mean and 1.96 standard deviations above the mean so that we have

$$(\overline{X} - 1.96\sigma_{\overline{X}}, \overline{X} + 1.96\sigma_{\overline{X}}) = (154.772, 156.198).$$

Notice that the standard deviation of the means of the heights corresponds to the algebraic standard error expression. In this case, the algebraic expression is much more efficient than the simulation. However, there are cases in which a simulation is more effective. For instance if we want to find the standard error of the median of a random variable that does not have a normal distribution.

Consider the number of texts received by twelve-year-old students. We can see in the density plot below that this variable does not follow a normal distribution and the center of the data is best represented by the median.



We see that the median number of text messages received yesterday from our sample of 1000 twelve-year-old students was 7 messages. Using 1000 simulations we see that the average of these medians was 7.219 text messages with a standard error of 1.015 messages. Using the simulation technique for estimating a 95% confidence interval generates an interval of (6,10), while the Central Limit Theorem method of estimating a 95% confidence interval generates an interval of (5.230, 9.208). Since the median of the population is likely to be a natural number, the interval of (6,10) is probably more appropriate for this setting.

16.2.2 Estimating Differences of Means

If we want to compare the heights of twelve-year-old students and seventeen-year-old students we can use the Census at School random sampler to generate a random sample of seventeen-year-old students from the United States. The average height of students in this sample is 171.105 cm. This seems to be significantly different from the average height of the twelve-year-old students, 155.485 cm, but it would be useful to have a way to test for such a difference. A standard process to determine if there is a difference between two means is to consider the random variable of the differences between the means of the samples, $(\overline{X_1} - \overline{X_2})$, which estimates the differences in the means of the populations, $(\mu_1 - \mu_2)$. The Central Limit Theorem implies that the standard error for this point estimator is $\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$.

Since the standard deviation of the height of the sample of twelve-year-old students is 11.520 and the standard deviation of the height of the sample of seventeen-year-old students is 12.489, we see that the standard error for the difference of means is

$$\sigma_{(\overline{X_1} - \overline{X_2})} = \sqrt{\frac{12.489^2}{1000} + \frac{11.520^2}{1000}} = 0.537$$

and so the 95% confidence interval for the difference in means is $(14.57, 16.67)$. Since zero is not inside of this confidence interval we can infer that the difference in heights is significant.

If we look at a sample of 1000 thirteen-year-old students we see that the average height of the 510 Females is 160.1176 and the average height of the 488 Males is 162.707. Using the standard deviations of the two samples we see that the standard error for the difference in means is

$$\sigma_{(\overline{X_1} - \overline{X_2})} = \sqrt{\frac{12.88^2}{488} + \frac{9.71^2}{510}} = 0.72$$

producing a 95% confidence interval for the difference between the means of (1.17, 4.01). A similar process for the 1000 twelve-year-old students estimates the average height of twelve-year-old females as 155.4825 ($n_1 = 485$) and males as 155.4874 ($n_2 = 515$) with a confidence interval for the difference in means of (-1.42, 1.43). So we can infer that there is no significant difference in the average heights of twelve-year-old students, but there is a significant difference in the average heights of thirteen-year-old students, based on gender.

Related Content Standards

- (HSS.IC.4) Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling.

16.2.3 Estimating Proportions

There are many times that we would like to find the proportion of the population that satisfies a certain condition. Election polling estimates the proportion of the population that will vote for certain candidates in an upcoming election. There are many times where a study wants to know the percentage of the population with certain demographic characteristics such as gender, race, or age.

If the proportion of the population that satisfies a certain condition is given by p, then in a random sample of size n we can label the outcome of the trial for each sample as Y_i whose value is a 0 if it does not satisfy the condition or a 1 if it satisfies the condition. So the expected value of the sample would be

$$E(X) = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} 1 \cdot p = np$$

since each of the cases is independent of the other cases. Therefore, we can estimate the proportion of the population with the proportion of the sample, $\hat{p} = \frac{X}{n}$.

Since the variance of a random variable, X, is the expected value of the squared deviation from the mean of X, $E(X)$,

$$\mathrm{Var}(X) = E\left((X - E(X))^2\right),$$

we can rewrite this expression as

$$\mathrm{Var}(X) = E\left((X - E(X))^2\right) = E\left(X^2 - 2XE(X) + E(X)^2\right) \tag{16.1}$$

$$= E(X^2) - 2E(X)E(X) + E(X)^2 = E(X^2) - E(X)^2. \tag{16.2}$$

So

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = E(X(X-1)) + E(X) - E(X)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

and the standard error of our estimator is

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}.$$

These estimators are appropriate once the binomial distribution approaches a normal distribution, with often occurs once both np and $n(p-1)$ are greater than 5.

From September 30 to October 3, 2020, Auburn University at Montgomery surveyed 1,072 registered voters and found that 397 people planned to vote for Biden and 611 people planned to vote for Trump in the 2020 Presidential election. We can then use simulations to create a confidence interval estimate for the percentage of the vote in Alabama that Donald Trump would earn.

```
cand = c(1,0)
px = c(0.57, 0.43)
nsims <- 1000 # number of simulations
m <- rep(NA,nsims) # empty vector to store means
for (i in 1:nsims){ m[i] <- mean(sample(cand, size=1072, replace=TRUE, prob=px))}
quantile(m,c(0.025,0.975))
```

This gives us a 95% confidence interval of $(0.54, 0.60)$ for the percentage of votes for Donald Trump in Alabama in 2020. This matches exactly the 95% confidence interval found using the standard error for the estimate,

$$\left(0.57 - 1.96 \cdot \sqrt{\frac{(0.57)(1-0.57)}{1072}}, 0.57 + 1.96 \cdot \sqrt{\frac{(0.57)(1-0.57)}{1072}}\right) = (0.54, 0.60).$$

16.2.4 Interpretations

The best way to understand a 95% confidence interval is that if one repeated the sampling procedure 100 times and created 100 95% confidence intervals, one would expect that 95 of these would contain the actual value of the parameter being estimated. So if we are comparing an estimate of a parameter for two samples and the 95% confidence intervals of these samples do not overlap we can be fairly confident the the values of the population parameters being estimated are distinct. If the confidence intervals overlap, we need to use other pieces of evidence to build an argument for the existence or non-existence of a difference between the two populations.

16.2.5 Exercises

1. In a study to measure the effects of texting while driving [He et al., 2014], 35 college-age participants were tested on their brake response times under different conditions. For the drive-only condition, the brake response times had a mean of 1.49 s with a standard deviation of 0.56 s. This leads to a 95% confidence interval of $(1.30, 1.68)$. A condition of driving while using a verbal texting system had a mean of 1.63 s with a standard deviation of 0.47 s, giving a 95% confidence interval of $(1.47, 1.79)$. The condition of driving and manually texting had a mean of 1.73, standard deviation of 0.48 s with a 95% confidence interval of $(1.57, 1.89)$.

    a. What conclusions can be drawn from these results and how would you provide evidence for your assertions?
    b. If you wanted to replicate this study with a stronger set of evidence, what sample size would be required? Explain your reasoning.

16.3 Hypothesis Testing

Following the order presented by Macdonald [1997] and Perezgonzalez [2015] we will introduce the historical development of two different views of hypothesis testing, with much of this material copied and adapted from Perezgonzalez [2015] under a Creative Commons license (CC BY).

16.3.1 Fisher's Approach

Although some steps in Fisher's approach may be worked out prior to the collect of data (e.g., the setting of hypotheses and levels of significance), the approach is eminently inferential and all steps can be set up once the research data are ready to be analyzed. Some of these steps can even be omitted in practice, as it is relatively easy for a reader to recreate them. Fisher's approach to data testing can be summarized in the five steps described below. We will analyze data collected from the Census at School project to study the difference in the heights of thirteen-year-old students based on gender. The analysis will be conducted with random samples of 15 females and 15 males to simulate the type of data that might be collected in a classroom.

Step 1–Select an appropriate test.

This step calls for selecting a test appropriate to the research goal of interest, with a consideration of properties of the variable considered. A list of some of the basic tests and their uses is given below.

Correlational Tests that look for an association between variables. + Pearson Correlation (ρ): Tests for the strength of association between two continuous variables + Kendall's Correlation (τ): Tests for the strength of association between two ordinal variables + Chi-Square (χ^2): Tests for the strength of association between two categorical variables.

Comparison of Means Tests that look for the difference between the means of variables + Paired t-test & Tests for the difference between two variables from the same population (e.g. a pre- and post-test score) + Independent t-test & Tests for the difference between two variables from different populations (e.g. comparing two subgroups based on gender)

In our example comparing the heights of male and female thirteen-year-old students from the United States we will be using an independent t-test since the variable of height is continuous and the two samples are independent from each other.

Step 2–Set up the null hypothesis (H_0).

The **null hypothesis** (H_0) derives naturally from the test selected in the form of an exact statistical hypothesis (e.g., $H_0 : \mu_1 - \mu_2 = 0$). The states that in the population of interest there is no change, no difference, or no relationship regarding a certain property of a parameter. It is called the null hypothesis because it stands to be nullified with research data.

Directional and non-directional hypotheses. With some research projects, the direction of the results is expected (e.g., one group will perform better than the other). In these cases, a directional null hypothesis covering all remaining possible results can be set (e.g., $H_0 : \mu_1 - \mu_2 \leq 0$). With other projects, however, the direction of the results is not predictable or of no research interest. In these cases, a non-directional hypothesis is most suitable (e.g., $H_0 : \mu_1 - \mu_2 = 0$).

For the example of comparing heights for thirteen-year-old students we will set the null hypothesis to be that there is no difference in the heights of males and females,

$$H_0 : \mu_{\text{males}} - \mu_{\text{females}} = 0.$$

Step 3-Calculate the theoretical probability of the results under H_0 (p).

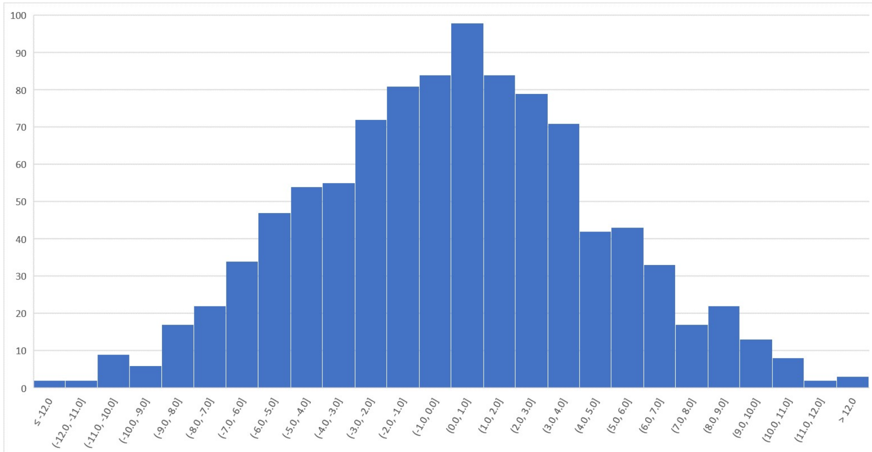
Once the corresponding theoretical distribution is established, the probability (p-value) of any datum under the null hypothesis is also established.

When the sample size, n, is small (generally less than 30) and the population standard deviation is unknown, we use a t-distribution as the theoretical distribution for the null hypothesis, usually with $n - 1$ degrees of freedom. For sample sizes larger than 30 we assume that the sample fits a normal distribution and we use the standard error as a substitute for the standard deviation of the population.

In our example of studying the heights of thirteen-year-old students we find the male heights sample mean (173.47) and standard deviation (16.81) and female heights sample mean (164.73) and standard deviation (3.53). This produces a standard error for the difference of means of 4.43. Using this as the standard deviation of the distribution for the null hypothesis we can generate the following 1000 samples with mean of 0, based on a normal distribution.

We see that data closer to the mean of the distribution, say between -5 and 5, have a greater probability of occurrence under the null distribution; that is, they appear more frequently and show a larger p-value (e.g., $p = 0.72$, or 720 times in a 1000 trials). On the other hand, data located further away from the mean have a lower probability of occurrence under the null distribution.

The p-value comprises the probability of the observed results and also of any other more extreme results (e.g., the probability of the actual difference between groups and any other difference more extreme than that). Thus, the p-value is a cumulative probability rather than an exact point probability: It covers the probability area extending from the observed results toward the tail of the distribution. For our example a difference of 8.73 has a smaller p-value (e.g., $p = 0.067$). This means that if there is no difference between the heights of male and female thirteen-year-old students and the heights of these students follow a t-distribution with a mean of zero, if we generate 1000 samples of the same size as the one used, 67 of these samples would have a average difference in heights the same or larger than the one generated in our sample.



Step 4–Assess the statistical significance of the results.

Fisher proposed tests of significance as a tool for identifying research results of interest, defined as those with a low probability of occurring as mere random variation of a null hypothesis. A research result with a low p-value may, thus, be taken as evidence against the null (i.e., as evidence that it may not explain those results satisfactorily). How small a result ought to be in order to be considered statistically significant is largely dependent on the researcher in question, and may vary from research to research. The decision can also be left to the reader, so reporting exact p-values is very informative.

Overall, however, the assessment of research results is largely made bound to a given level of significance, by comparing whether the research p-value is smaller than such level of significance or not:

- If the p-value is approximately equal to or smaller than the level of significance, the result is considered statistically significant.
- If the p-value is larger than the level of significance, the result is considered statistically non-significant.

Level of significance. The level of significance is a theoretical p-value used as a point of reference to help identify statistically significant results. There is no need to set up a level of significance a priori nor for a particular level of significance to be used in all occasions, although levels of significance such as 5% or 1% may be used for convenience. This highlights an important property of Fisher's levels of significance: They do not need to be rigid (e.g., p-values such as 0.049 and 0.051 have about the same statistical significance around a convenient level of significance of 5%).

Another property of tests of significance is that the observed p-value is taken as evidence against the null hypothesis, so that the smaller the p-value the stronger the evidence it provides. This means that it is plausible to gradate the strength of such evidence with smaller levels of significance. For example, if using 5% as a convenient level for identifying results which are just significant, then 1% may be used as a convenient level for identifying highly significant results and 0.1% for identifying extremely significant results.

Step 5–Interpret the statistical significance of the results.

A significant result is literally interpreted as a dual statement: Either a rare result that occurs only with probability p (or lower) just happened, or the null hypothesis does not explain the research results satisfactorily. Such literal interpretation is rarely encountered, however, and most common interpretations are in the line of "The null hypothesis did not seem to explain the research results well, thus we inferred that other processes—which we believe to be our experimental manipulation—exist that account for the results," or "The research results were statistically significant, thus we inferred that the treatment used accounted for such difference."

Non-significant results may be ignored, although they can still provide useful information, such as whether results were in the expected direction and about their magnitude. In fact, although always denying that the null hypothesis could ever be supported or established, Fisher conceded that non-significant results might be used for confirming or strengthening it.

In our case of comparing heights of male and female thirteen-year-old students we found a difference in means of 8.73 cm with a p-value of 0.067. While this does not meet the standard threshold of 0.05, we can still argue that we should reject the null hypothesis that male students are the same height as female students. We may want to replicate this study with a new experiment with a larger sample size. By increasing the sample size, it is much more likely to obtain a difference that is statistically significant.

Highlights of Fisher's approach

Flexibility. Because most of the work is done a posteriori, Fisher's approach is quite flexible, allowing for any number of tests to be carried out and, therefore, any number of null hypotheses to be tested (a correction of the level of significance may be appropriate).

Better suited for ad-hoc research projects. Given above flexibility, Fisher's approach is well suited for single, ad-hoc, research projects, as well as for exploratory research.

Inferential. Fisher's procedure is largely inferential, from the sample to the population of reference, albeit of limited reach, mainly restricted to populations that share parameters similar to those estimated from the sample.

No alternative hypothesis. One of the main critiques to Fisher's approach is the lack of an explicit alternative hypothesis, because there is no point in rejecting a null hypothesis without an alternative explanation being available. However, Fisher considered alternative hypotheses implicitly—these being the negation of the null hypotheses—so much so that for him the main task of the researcher—and a definition of a research project well done—was to systematically reject with enough evidence the corresponding null hypothesis.

16.3.2 Neyman and Pearson's Approach

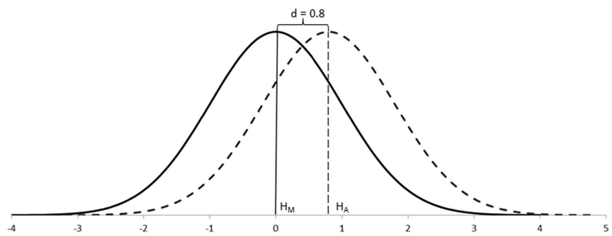
Jerzy Neyman and Egon Sharpe Pearson developed an alternative approach to data testing that is more mathematical than Fisher's and does much of its work at the planning stage of the research project [Macdonald, 1997]. It introduces a number of constructs, some of which are similar to those of Fisher and is summarized in the following eight main steps.

During the planning stages.

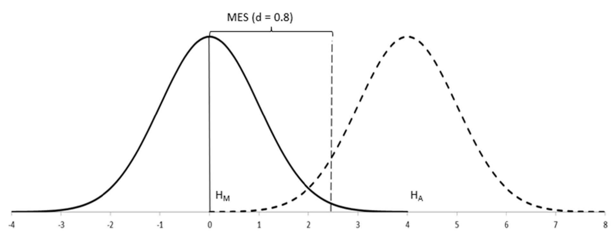
Step 1–Set up the expected effect size in the population.

The main conceptual innovation of the Neyman-Pearson approach is the consideration of explicit alternative hypotheses when testing research data. In their simplest postulate, the alternative hypothesis represents a second population that sits alongside the population of the main hypothesis on the same continuum of values. These two groups differ by some degree: the **effect size**.

For example, Cohen's [1988] conventions for capturing differences between groups, d, were based on the degree of visibility of such differences in the population: the smaller the effect size, the more difficult to appreciate such differences; the larger the effect size, the easier to appreciate such differences. Thus, effect sizes also double as a measure of importance in the real world.



When testing data about samples, however, statistics do not work with unknown population distributions but with distributions of samples, which have narrower standard errors. In these cases, the effect size can still be defined as above because the means of the populations remain unaffected, but the sampling distributions would appear separated rather than overlapping. Because we rarely know the parameters of populations, it is their equivalent effect size measures in the context of sampling distributions which are of interest.



As we shall see below, the alternative hypothesis is the one that provides information about the effect size to be expected. However, because this hypothesis is not tested, the Neyman-Pearson approach largely ignores its distribution except for a small percentage of it, which is called β. Therefore, it is easier to understand the approach if we peg the effect size to beta and call it the expected **minimum effect size (MES)**. The minimum effect size effectively represents that part of the main hypothesis that is not going to be rejected by the test (i.e., MES captures values of no research interest which you want to leave under the main hypothesis, H_M).

Step 2–Select an optimal test.

While the Neyman-Pearson approach allows for differentiation between tests described as the power of the test, the most common tests used are the same as those in the Fisher approach.

Step 3–Set up the main hypothesis (H_M).

The Neyman-Pearson approach considers, at least, two competing hypotheses, although it only tests data under one of them. The hypothesis which is the most important for the research (i.e., the one you do not want to reject too often) is the one tested. This hypothesis is better off written so as to incorporate the minimum expected effect size within its postulate (e.g., $H_M : \mu_1 - \mu_2 \in (0 - \text{MES}, 0 + \text{MES})$), so that it is clear that values within such minimum threshold are considered reasonably probable under the main hypothesis, while values outside such minimum threshold are considered as more probable under the alternative hypothesis.

The main aspect to consider when setting the main hypothesis is the Type I error you want to control for during the research.

Definition 16.1. A **Type I** error (or error of the first class) is made every time the main hypothesis is wrongly rejected.

The **alpha level,** α**,** of a a hypothesis test is the probability that the test will lead to a Type I error.

Because the hypothesis under test is your main hypothesis, this is an error that you want to minimize as much as possible. Neyman and Pearson often worked with convenient alpha levels such as 5% ($\alpha = 0.05$) and 1% ($\alpha = 0.01$), although different levels can also be set. The main hypothesis can, thus, be written so as for incorporating the alpha level in its postulate (e.g., $H_m : \mu_1 - \mu_2 \in (0 - \text{MES}, 0 + \text{MES}), \alpha = 0.05$), to be read as the probability level at which the main hypothesis will be rejected in favor of the alternative hypothesis.

Step 4–Set up the alternative hypothesis (H_A).

One of the main innovations of the Neyman-Pearson approach is the consideration of alternative hypotheses. The alternative hypothesis is written so as for incorporating the minimum effect size within its postulate (e.g., $H_A : \mu_1 - \mu_2 \notin (0 - \text{MES}, 0 + \text{MES})$). This way it is clear that values beyond such minimum effect size are the ones considered of research importance.

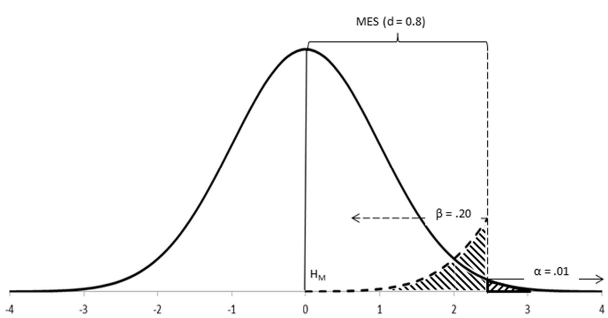
Among things to consider when setting the alternative hypothesis are the expected effect size in the population (see above) and the Type II error you are prepared to commit.

Definition 16.2. A **Type II** error (or error of the second class) is made every time the main hypothesis is wrongly retained (thus, every time H_A is wrongly rejected).

Beta (β**)** is the probability of committing a Type II error in the long run and is, therefore, a parameter of the alternative hypothesis

Making a Type II error is less critical than making a Type I error, yet you still want to minimize the probability of making this error once you have decided which alpha level to use.

You want to make beta as small as possible, although not smaller than alpha (if β needed to be smaller than α, then H_A should be your main hypothesis, instead). Neyman and Pearson proposed 20% ($\beta = 0.20$) as an upper ceiling for beta, and the value of alpha ($\beta = \alpha$) as its lower floor. For symmetry with the main hypothesis, the alternative hypothesis can, thus, be written so as for incorporating the beta level in its postulate (e.g., $H_A : \mu_1 - \mu_2 \neq 0 \pm \text{MES}, \beta = 0.20$).



Step 5–Calculate the sample size (N) required for good power ($1 - \beta$).

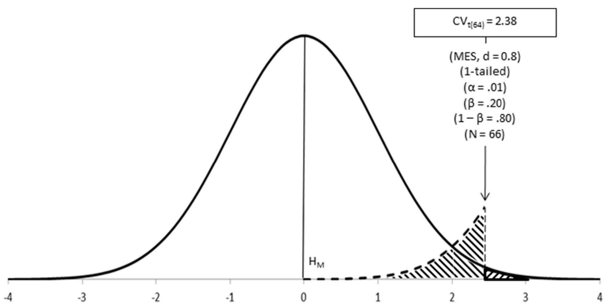
Neyman-Pearson's approach is predominantly dependent upon the process of designing the experiment or observation in order to ensure that the research to be done has good power. Power is the probability of correctly rejecting the main hypothesis in favor of the alternative hypothesis (i.e., of correctly accepting H_A). It is the mathematical opposite of the Type II error (thus, $1 - \beta$).

Step 6–Calculate the critical value of the test ($\text{CV}_{\textbf{test}}$, or $\text{Test}_{\textbf{crit}}$).

Some of above parameters (α and N) can be used for calculating the critical value of the test; that is, the value to be used as the cut-off point for deciding between hypotheses.

Table 16.1:

	H_m True	H_m False
Retain H_m	Decision Correct	Type II Error (probability β)
Reject H_m	Type I Error (probability α)	Decision Correct (power $1-\beta$)



After the data is available for analysis.

Step 7–Calculate the test value for the research.

In order to carry out the test, some unknown parameters of the populations are estimated from the sample (e.g., variance), while other parameters are deduced theoretically (e.g., the distribution of frequencies under a particular statistical distribution). The statistical distribution so established thus represents the random variability that is theoretically expected for a statistical main hypothesis given a particular research sample, and provides information about the values expected at different locations under such distribution.

Step 8–Decide in favor of either the main or the alternative hypothesis.

Neyman-Pearson's approach is rather mechanical once the design steps have been satisfied. Thus, the analysis is carried out as per the optimal test selected and the interpretation of results is informed by the mathematics of the test, following on the design set up for deciding between hypotheses:

- If the observed result falls within the critical region, reject the main hypothesis and accept the alternative hypothesis.
- If the observed result falls outside the critical region and the test has good power, accept the main hypothesis.
- If the observed result falls outside the critical region and the test has low power, conclude nothing.

16.3.2.1 Highlights of Neyman-Pearson's Approach

More powerful. Neyman-Pearson's approach is more powerful than Fisher's for testing data in the long run. However, repeated sampling is rare in research.

Better suited for repeated sampling projects. Because of above, Neyman-Pearson's approach is well-suited for repeated sampling research using the same population and tests, such as industrial quality control or large scale diagnostic testing.

Deductive. The approach is deductive and rather mechanical once the design steps have been set up.

Less flexible than Fisher's approach. Because most of the work is done in the design stage, this approach is less flexible for accommodating tests not thought of beforehand and for doing exploratory research.

Defaults easily to Fisher's approach. As this approach looks superficially similar to Fisher's, it is easy to confuse both and forget what makes Neyman-Pearson's approach unique. If the information provided by the alternative hypothesis, MES and β, is not taken into account for designing research with good power, data analysis defaults to Fisher's test of significance.

16.3.3 Exercises

1. A statistical analysis of a quantitative value of two groups of individuals is carried out and a test of the hypothesis that there is no difference between the two groups concerning that variable. The hypothesis test results in a p-value of 0.01. What would be an appropriate interpretation of this result?

2. The results of a hypothesis test are statistically significant for a significance level of $\alpha = 0.05$. What does this mean?

Bibliography

Deborah Loewenberg Ball, Mark Hoover Thames, and Geoffrey Phelps. Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5):389–407, 2008.

Anna Bargagliotti, Christine Franklin, Pip Arnold, Rob Gould, Sheri Johnson, Leticia Perez, and Denise A. Spangler. *Pre-K–12 Guidelines for Assessment and Instruction in Statistics Education II (GAISE II)*. American Statistical Association, Alexandria, VA, 2020.

John T. Behrens and Chong-ho Yu. Exploratory data analysis. In John A. Schinka and Wayne F. Velicer, editors, *Research Methods in Psychology*, volume 2 of *Handbook of Psychology*, chapter 2. John Wiley & Sons, Inc., 2003.

Joseph Bertrand. *Traité D'Arithmétique*. L. Hachette & Co., Paris, 1840. URL https://gallica.bnf.fr/ark:/12148/bpt6k77735p.

George Boole. *The Mathematical Analysis of Logic*. Macmillan, Barclay, & Macmillan, Cambridge, 1847.

Carl B. Boyer and Uta C. Merzbach. *A History of Mathematics*. John Wiley & Sons, Inc., 2 edition, 1991.

Georg Cantor. Ueber eine elementare frage der mannigfaltigkeitslehre. *Jahresbericht der Deutschen Mathematiker-Verreinigung*, 1890/91:75–78, 1891. URL https://www.digizeitschriften.de/dms/img/?PID=GDZPPN002113910.

Augustin-Louis Cauchy. *Cours D'Analyse de L'École Royale Polytechnique*, volume I. Analyse Algébrique. de l'Imprimerie Royale, Paris, 1821.

Arthur Cayley. On the theory of groups, as depending on the symbolic equation $\theta^n = 1$. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 7(42):40–47, 1854.

Beth L. Chance. Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10(3), 2002. URL http://jse.amstat.org/v10n3/chance.html.

George W. Cobb and David S. Moore. Mathematics, statistics, and teaching. *the American Mathematical Monthly*, 104(9):801–823, 1997. URL https://doi.org/10.1080/00029890.1997.11990723.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Psychology Press, New York, NY, 2 edition, 1988.

Conference Board of the Mathematical Sciences. *The Mathematical Education of Teachers*. American Mathematical Society and Mathematical Association of America, Providence RI and Washington DC, 2001.

Conference Board of the Mathematical Sciences. *The Mathematical Education of Teachers II*. American Mathematical Society and Mathematical Association of America, Providence RI and Washington DC, 2012.

Roger Cotes. *Harmonia Mensurarum*. Cambridge, 1722. Edited by Robert Smith.

Augustus De Morgan. *Formal Logic; or, The Calculus of inference, necessary and probable*. Taylor and Walton, London, 1847.

Richard Dedekind. *Stetigkeit und irrationale Zahlen*. Friedrich Bieweg und Gohn, Braunschweig, 1872. URL https://archive.org/details/stetigkeitundir00dedegoog.

Barbara Edwards and Michael B. Ward. The role of mathematical definitions in mathematics and in undergraduate mathematics courses. In Marilyn Carlson and Chris Rasmussen, editors, *Making the Connection: Research and Teaching in Undergraduate Mathematics Education*, chapter 17, pages 223–232. Mathematical Association of America, 2008.

Leonhard Euler. *Introductio in Analysin Infinitorum*, volume 1. Marcus-Michael Bousquet & Co., Lausanne, 1748. URL https://gallica.bnf.fr/ark:/12148/bpt6k33510/.

William Ewald and Willfried Sieg. Lectures on the infinite. In *David Hilbert's Lectures on the Foundations of Arithmetic and Logic, 1917-1933*. Springer, 2013.

Christine Franklin, Gary D. Kader, Denise A. Mewborn, J. Moreno, R. Peck, M. Perry, and Richard L. Schaeffer. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*. American Statistical Association, Alexandra, VA, 2007.

Craig G. Fraser, Wilbur R. Knorr, John L. Berggren, Menso Folkerts, and Jeremy John Gray. Mathematics. In *Encyclopædia Britannica*. Encyclopædia Britannica, inc., May 2019. URL https://www.britannica.com/science/mathematics.

Sol Garfunkel and Michelle Montgomery, editors. *GAIMME: Guidelines for Assessment and Instruction in Mathematical Modeling Education*, Philadelphia, PA, 2019. Consortium for Mathematics and Its Applications (COMAP) and Society for Industrial and Applied Mathematics (SIAM). URL https://www.siam.org/Publications/Reports/Detail/guidelines-for-assessment-and-instruction-in-mathematical-modeling-education. 2nd edition.

Alexander Osipovich Gel'fond. Sur le septiéme probléme de Hilbert. *Belletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, 4:623–634, 1934.

Camilla Gilmore, Silke M. Göbel, and Matthew Inglis. *An Introduction to Mathematical Cognition*. International Texts in Developmental Psychology. Taylor & Francis, 2018. ISBN 9781317410102. URL https://books.google.com/books?id=cVZgDwAAQBAJ.

J. He, A. Chaparro, B. Nguyen, R.J. Burge, J. Crandall, B. Chaparro, R. Ni, and S. Cao. Texting while driving: Is speech-based text entry less risky than handheld text entry? *Accident Analysis & Prevention*, 72:287 – 295, 2014. ISSN 0001-4575. doi: https://doi.org/10.1016/j.aap.2014.07.014. URL http://www.sciencedirect.com/science/article/pii/S0001457514002176.

T. L. Heath. *The Thirteen Books of Euclid's Elements: Translated from the Text of Heiberg with Introduction and Commentary*, volume 1, Introduction and Books I and II. University Press, Cambridge, 1908a.

T. L. Heath. *The Thirteen Books of Euclid's Elements: Translated from the Text of Heiberg with Introduction and Commentary*, volume 2, Books III-IX. University Press, Cambridge, 1908b.

E. Heine. Die elemente der funtionenlehre. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*, 74:172–188, 1872.

David Hilbert. *The Foundations of Geometry*. The Open Court Publishing Company, Chicago, second edition, 1910. Authorized Translation by E. J. Townsend.

David Klein. A brief history of American K-12 mathematics education in the 20th century. In James M. Royer, editor, *Mathematical Cognition*, pages 175–225. Information Age Pub., 2003.

Felix Klein. *Vergleichende Betrachtungen uber neuere geometrische Forschungen*. Andreas Deichart, Erlangen, 1872.

Jean Leurechon. *SelectæPropositiones in Tota Sparsim Mathematica Pulcherrimæ*. Mussiponti, Lotharingiæ, 1629. URL https://books.google.com/books?id=RGwTAAAAQAAJ.

Ronald R. Macdonald. On statistical testing in psychology. *British Journal of Psychology*, 88(2), 1997. doi: 10.1111/j.2044-8295.1997.tb02638.x.

National Commission on Excellence in Education. *A Nation at Risk: The Imperative for Educational Reform.* Department of Education, Washington, D.C., 1983.

NCTM. *An Agenda for Action: Recommendations for School Mathematics of the 1980s.* National Council of Teachers of Mathematics, Reston, VA, 1980.

NCTM. *Curriculum and Evaluation Standards for School Mathematics.* National Council of Teachers of Mathematics, Reston, VA, 1989.

NCTM. *Professional Teaching Standards.* National Council of Teachers of Mathematics, Reston, VA, 1991.

NCTM. *Assessment Standards.* National Council of Teachers of Mathematics, Reston, VA, 1995.

NCTM. *Principles and Standards for School Mathematics.* National Council of Teachers of Mathematics, Reston, VA, 2000.

NCTM. *Standards for the Preparation of Secondary Mathematics Teachers.* National Council of Teachers of Mathematics, Reston, VA, 2020.

NGA-CCSSO. *Common Core State Standards Mathematics.* National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C., 2010.

Jose D. Perezgonzalez. Fisher, Neyman-Pearson or NHST? a tutorial for teaching data testing. *Frontiers in Psychology*, 6(223), 2015. doi: 10.3389/fpsyg.2015.00223.

G. Polya. *How to Solve It: A new aspect of mathematical method.* Doubleday & Company, Inc., Garden City, NY, 2 edition, 1957.

João Pedro Ponte. The history of the concept of function and some educational implications. *The Mathematics Educator*, 3(2):3–8, 1992.

Ralph A. Raimi and Lawrence S. Braden. State mathematics standards: An appraisal of math standards in 46 States, the District of Columbia, and Japan. Technical report, Thomas B. Fordham Foundation, Washington, DC, 1998.

Barbara Reys and Glenda Lappan. Consensus or confusion? the intended math curriculum in State-level standards. *Phi Delta Kappan*, 88(9):676–680, May 2007.

Walter Rudin. *Principles of Mathematical Analysis.* McGraw-Hill, Inc., 3 edition, 1976.

Laney Sammons. *Teaching Students to Communicate Mathematically.* ASCD, Alexandria, VA, 2018.

Theodor Schneider. Transzendenzuntersuchungen periodischer funktionen I. transzendenz von potenzen. *Journal für die reine und angewandte Mathematik (Crelle's Journal)*, 172:65–69, 1935.

School Mathematics Study Group. *Mathematics for High School: Geometry.* Yale University, New Haven, 1960.

Lee S Shulman. Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2):4–14, 1986.

H. A. Thurston. *The Number-System.* Blackie & Son, Limited, London, 1956. URL https://archive.org/details/TheNumberSystem.

UNESCO/UNODC. *Empowering students for just societies: a handbook for secondary school teachers.* UNESCO/UNODC, 2019.

Zalman Usiskin and Jennifer Griffin. *The Classification of Quadrilaterals: A Study in Definition.* Research in Mathematics Education. Information Age Publishing Inc., Charlotte, 2008. ISBN 9781607526001.

Zalman Usiskin, Anthony Peressini, Elena Anne Marchisotto, and Dick Stanley. *Mathematics for High School Teachers: An advanced perspective*. Pearson Education, Inc., Upper Saddle River, NJ, 2003.

J. Van Dormolen and O. Zaslavsky. The many facets of a definition: The case of periodicity. *Journal of Mathematical Behavior*, 22:91–196, 2003.

Shlomo Vinner. The role of definitions in the teaching and learning of mathematics. In D. Tall, editor, *Advanced Mathematical Thinking*, chapter 5, pages 65–80. Kluwer Academic Press, Dordrecht, 1991.

John von Neumann. Zur einführung der transfiniten zahlen. *Acta Litterarum ac Scientiarum RegiæUniversitatis HungaricæFrancisco-Josephinae, Sectio Scientiarum Mathematicarum*, 1:199–208, 1923.

Alfred North Whitehead and Bertrand Russell. *Principia Mathematica*, volume 1. Cambridge University Press, 1910.

Alfred North Whitehead and Bertrand Russell. *Principia Mathematica*, volume 2. Cambridge University Press, 1912.

Alfred North Whitehead and Bertrand Russell. *Principia Mathematica*, volume 3. Cambridge University Press, 1913.