2018–2019

MATHEMATICS

Resource Guide

Overview of Permutations and Combinations, Algebra, and Statistics

USAD® United States Academic Decathlon®

The vision of the United States Academic Decathlon® is to provide all students the opportunity to excel academically through team competition.

Toll Free: 866-511-USAD (8723) • **Direct:** 712-326-9589 • **Fax:** 651-389-9144 • **Email:** info@usad.org • **Website:** www.usad.org

Table of Contents

Introduction

In this year's *Mathematics Resource Guide*, we will study three connected areas of mathematics: permutations and combinations, algebra, and statistics. The idea that connections exist between these areas of mathematics may seem strange to you now, but connections and relationships between seemingly disconnected topics such as these are at the heart of mathematics. It is our hope that after reading through this *Mathematics Resource Guide*, you will be aware of and interested in these connections among different areas of mathematics.

The first section of the resource guide will cover permutations and combinations—topics that are sometimes discussed in a general high school math sequence, but perhaps are only given a cursory treatment that often leaves students without a good sense of their power, flexibility, and array of applications. Combinations in particular are extremely important and are used in a wide variety of mathematical contexts, and so a comfort level with these mathematical structures will pay dividends in your future study of mathematics.

Section 2 will focus on the topic of algebra. Algebra is a term that has seemingly become synonymous with high school mathematics, and yet the algebra most students learn in high school mathematics courses is only a small fraction of the field of algebra. Our purpose in this section is to highlight some important algebraic ideas and patterns that are often overlooked in a standard study of high school algebra. Being comfortable with arithmetic and geometric sequences and series is at least as important as knowing how to factor a quadratic, but most students spend comparatively little time looking at these sequences and series. Sigma notation for series becomes increasingly important in the study of calculus and mathematics beyond calculus, so this resource guide aims to provide you with plenty of opportunities to practice reading and manipulating sigma notation. The Binomial Expansion Theorem is another foundational topic often left out or deemphasized in traditional high school mathematics, and we will use our previous work with combinations to make sense of this important theorem. Finally, Euler's constant, e, is often used but rarely properly understood, and so we will look at e both from a contextual and a mathematical perspective.

Statistics is a branch of mathematics often only briefly covered in high school mathematics and misunderstood by much of the general public. In the final section of the *Mathematics Resource Guide*, we will take a more mathematical look at some of the foundations of statistics, and discuss the reasons different statistical measures, such as mean, median, variance, and standard deviation, were developed. The foundational concepts of probability distributions, the Binomial Distribution, and the Normal Distribution will also be investigated.

We hope you find this year's *Mathematics Resource Guide* interesting and insightful as we look at connections between these seemingly disparate areas of mathematics. Enjoy!

Section 1

Overview of Permutations and Combinations

In many high school mathematics sequences, permutations and combinations are often left out all together. When discussed, they are not usually given sufficient time for students to develop an appreciation and mastery of these topics. This is certainly not the fault of high school mathematics teachers—there is simply too much good content to discuss and not enough time! Yet permutations and combinations are extremely important concepts that underpin a great deal of higher mathematics. Indeed, one of the fields of current mathematical study and research is called *combinatorics*.

Studying permutations and combinations also has great benefits for high school students. Initially these topics are fairly accessible, and some of the early problems can be approached quickly and solved easily. Yet, despite their humble beginnings, they eventually become powerful tools for solving a wide array of problems, and this flexibility makes them very useful. The study of permutations and combinations encourages and requires visualization, algebraic fluency, and an attention to accurate calculations—all positive mathematical traits.

In last year's *Mathematics Resource Guide*, some time was spent discussing permutations and combinations and relating them to probability. As we will use permutations and combinations again in this year's *Mathematics Resource Guide*, we will begin with a refresher on these concepts. If you are able, you can review the portions of last year's *Mathematics Resource Guide* on these topics. If not, have no fear! Everything you will need to know and understand about permutations and combinations will be discussed in this year's resource guide.

1.1 THE MULTIPLICATION PRINCIPLE

We will begin with a familiar idea from elementary school mathematics: multiplication. Multiplication is most often useful in situations of repeated addition: rather than add up $3 + 3 + 3 + 3 + 3 + 3 + 3 + 3 + 3$, which could become long and tedious, we instead use multiplication and write $3 \cdot 9 = 27$. Many times (hah!) multiplication is introduced *as* repeated addition, and if asked, many people will give "repeated addition" as a definition, or *the* definition, of multiplication. (It turns out some mathematicians may disagree with this definition, but that is another story.) Multiplication is in some ways the first important mathematical notation encountered, as it represents an operation that is a shortening of a (potentially) long process of calculation. Let's start with some straightforward examples.

EXAMPLE 1.1A: Toni is having a birthday party next week and wants to give everyone six pieces of candy in their gift bag. If Toni is inviting 15 people to the birthday party, how many pieces of candy does Toni need?

SOLUTION: One way to solve this problem is to add up six 15 times: $6 + 6 + 6 + 6 + 6 + 6 + 6 + 6 + 6 + 6 + 6 + 6 + 6 + 6 + 6$. This makes sense because each of the 15 guests gets six pieces of candy. Although we might expect a first- or second-grader to solve the problem using addition, we are slightly more mathematically sophisticated, and say $15 \cdot 6 =$ **90**.

EXAMPLE 1.1B: Toni goes to the store to purchase the candy. Unfortunately, the only candy the store has is little packages of candies with 7 pieces of candy in each package. Toni decides to give each guest six packages of this candy. How many pieces of candy does Toni purchase in order to fill the gift bags?

SOLUTION: Toni was planning on getting 90 pieces of candy, since $15 \cdot 6 = 90$. But now each of the 90 packages contains 7 pieces of candy, so Toni will purchase $15 \cdot 6 \cdot 7 =$ **630** pieces of candy.

Example 1.1c: Once at the party, everyone decides they need to open their gift bag immediately after receiving it, rather than wait until they get home. It turns out the individual pieces of candy inside the packages were a bit larger than Toni thought, and they need to be eaten in two bites each. Assuming every guest does this uniformly, how many bites will it take for all fifteen guests to eat all of the candy?

Solution: Each of the 630 pieces of candy will take two bites to eat, so there will be a total of $15 \cdot 6 \cdot 7 \cdot 2 =$ **1,260** bites before all the candy is eaten.

Although this a nice series of problems, and it certainly represents a good use of multiplication, it doesn't seem that this solution strategy generalizes easily. Each guest has six packages of candy, and each package of candy has seven pieces, and each piece takes two bites to eat. How often do situations like this really arise?

Let's take a step back for a moment and imagine what we would do to solve this problem if we didn't understand multiplication very well, or hadn't learned multiplication yet. One option would be to make a list of all bites needed to eat the candy. We could make up names for each of the fifteen guests, assign the packages of candy numbers, number off the pieces of candy, and start writing:

Amy, Package 1, Candy 1, Bite 1
Amy, Package 1, Candy 1, Bite 2
Amy, Package 1, Candy 2, Bite 1
Amy, Package 1, Candy 2, Bite 2

As you can imagine, this would take quite a while. After we get through all of Amy's seven candies in Package 1, we would move to Package 2, and only after we write out all of Amy's six packages could we even start on Bob's candies.

While writing out the entire list (with 1,260 entries!) would clearly be awful, thinking about the giant list is helpful. We can see there are 15 different possibilities for the person, 6 different possibilities for the package, 7 different possibilities for the candy, and 2 different possibilities for the bite. This mathematical structure—keeping track of the different possibilities for each entry in a list—*is* more applicable and can be applied to a variety of contexts.

EXAMPLE 1.1D: Many license plates have three letters A–Z followed by three numbers 0–9. Assuming there are no restrictions on the letters or numbers used, how many different license plates are possible?

SOLUTION: Again, we clearly don't want to write out the entire list of possibilities, but imagining those possibilities is a good place to start. Our list would begin with the license plate *AAA 000*, followed by *AAA 001*, and *AAA 002*. Eventually we would get to *AAA 999*, followed by *AAB 000* (exciting!). Only after a brutally long time would we ever get to start a license plate with B, let alone get all the way down to the end of the list at *ZZZ 999*. So how many license plates are possible?

With 26 choices for each of the letters and 10 choices for each of the numbers, there will be $26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 =$ **17,576,000** different possible license plates. (Good thing we didn't try to write them all out!)

Keeping these examples in mind, it seems we are ready to generalize.

THE MULTIPLICATION PRINCIPLE

When listing out all the possibilities for k items, the total number of entries in this list is given by $n_1 \cdot n_2 \cdot n_3 \ldots n_k$, where n_k is the number of possibilities for the k^{th} item. For example, n_3 is the number of possibilities for the third item.

In terms of the candy for Toni's party, there were four items on the list: the person, the package, the piece of candy, and the bite. Since there were 15, 6, 7, and 2 choices, respectively, for each of these items, the answer was $15 \cdot 6 \cdot 7 \cdot 2$.

Let's look at one more example before we make this any more complicated. Be sure to think through *why* the multiplication principle makes sense; memorizing and imitating formulas doesn't help when those formulas need to be modified to apply them to new situations.

EXAMPLE 1.1E: Every morning, Jesse makes a sandwich to take to school for lunch. Jesse likes sandwiches with bread, cheese, lunchmeat, and a dressing. At the local grocery store, there are three choices for bread (white, wheat, and whole grain), four choices for cheese (cheddar, Colby jack, mozzarella, and Swiss), seven choices for lunchmeat (turkey, smoked turkey, ham, honey ham, roast beef, salami, and pastrami), and three choices for dressing (mayonnaise, mustard, and spicy mustard). How many different sandwiches can Jesse make?

SOLUTION: The Multiplication Principle says the answer is $3 \cdot 4 \cdot 7 \cdot 3 =$ **252**, but let's make sure we understand why this is the case. If our sandwich consisted of only bread, there would be three different sandwiches possible. Including cheese, each type of bread can be paired with four different cheeses, making a total of 12 possible sandwiches. Each of these 12 bread-cheese sandwiches can have one of seven lunchmeats added, bringing us to 84 sandwiches. And each of these 84 bread-cheese-lunchmeat sandwiches can have one of three dressings, for a total of **252** possible sandwiches.

1.2 PERMUTATIONS

The problems we have considered thus far are nice because each category is distinct, and there is no real possibility for confusion about which category contains a given object. Usually, we would not think *cheddar cheese, turkey, ham, mustard* would be a legitimate sandwich in the example above. When categories potentially overlap, or when we are repeatedly picking objects from the same category, we need to be more careful.

Let's consider an example. At the beginning of math class every day, Mr. Smith selects students to write up homework problems on the board. These problems are then discussed as a class. There are 26 students in Mr. Smith's math class, and he randomly selects a student to write up each of the first five problems. How many different ways can students be assigned to the problems?

This problem is more complicated than the sandwich or candy problems because we are selecting objects from the same group each time. Mr. Smith is not selecting students from five different classes or from different groups within his class. In this way, the problem is more like the license plate problem. In that problem, we were selecting repeatedly from two different groups. Here we are repeatedly selecting objects from only one group, the class of 26 students.

The first thing we have to determine is if any student can be selected more than once. In our imaginary list of all possible arrangements, is *Fred, Fred, Fred, Fred, Fred* an allowable selection? Or if Fred is selected to write up the first problem, is he "safe" from writing up problems #2 – 5?

This is an important distinction and can drastically impact the answer to the question. Fortunately, this does not significantly alter the way we think about the problem, just the answer we get. Hopefully whether or not objects can be repeated is clear from the context and description of the problem. Sometimes, the phrases "with replacement" or "without replacement" are used to clarify whether or not an object from a category can be selected more than once. This example will be presented for you to solve in the review problems at the end of this section. Let's now work through a few examples here.

EXAMPLE 1.2A: How many ways can three different cards be drawn with replacement from a standard 52-card deck?

SOLUTION: Since the cards are drawn with replacement, each card is put back into the deck before the next card is drawn. So, the same card could be drawn all three times, and there are 52 choices for the first card, 52 choices for the second card, and 52 choices for the third card. Therefore, there are $52 \cdot 52 \cdot 52 =$ **140,608** different ways three cards can be drawn with replacement.

EXAMPLE 1.2B: How many ways can three different cards be drawn without replacement from a standard 52-card deck, assuming the order of drawing cards matters?

SOLUTION: Since the cards are drawn without replacement, each card is kept when the next card is drawn. This means there are 52 choices for the first card, but only 51 choices for the second card, and 50 choices for the third card. Therefore, there are $52 \cdot 51 \cdot 50 =$ **132,600** different ways three different cards can be drawn without replacement.

There are two important things to consider from this pair of problems. The first is the difference in the answers. Although there are more ways to draw three cards with replacement than without, this difference is perhaps not as large as expected. Because there are so many cards to choose from, and we are selecting so few of them, most of the entries in our imaginary list of all 140,608 ways to write out three cards with replacement will contain three different cards. Therefore, the vast majority of these entries will be included in our imaginary list of all 132,600 ways to select three different cards. It seems as we select more and more cards from the deck, this difference should become greater.

The second idea that should be considered carefully is that when drawing is done without replacement, the order in which the cards are drawn makes a difference. Maybe the cards are being drawn and then placed face-up to create some form of "lineup," but most of the time when cards are drawn from a deck, the order in which they are dealt doesn't matter. How can this potentially more realistic situation be dealt with (hah!)? We will return to this line of questioning in a moment. For now, let's consider the difference between selecting with replacement and without replacement (assuming order matters).

EXAMPLE 1.2C: How many different ways can 52 cards be drawn from a deck with replacement (assuming the order of selection matters)?

SOLUTION: Since there are 52 choices for the first card, and 52 choices for the second card, and the third card, etc., we could write $52 \cdot 52 \cdot 52 \ldots 52$, with fifty-two 52's in this list. Rather than write out the numerical value of this calculation (which contains 90 digits), we use mathematical notation to say the answer is $\mathbf{52^{52}}$. Not only is this a nice way to write the answer, it also gives us some insight into the problem, as we can see the answer is fifty-two 52's all multiplied together.

EXAMPLE 1.2D: How many different ways can 52 cards be drawn from a deck without replacement (where the order of the cards matters)?

SOLUTION: Notice that eventually every card in the deck will be drawn. This means that when it is time to draw the last card, there is only one option, since every other card has already been drawn. So, there are 52 choices for the first card, but then 51 choices for the second card, and 50 choices for the third card, and so on down to only 1 choice for the last card. Therefore, the calculation we need to perform is $52 \cdot 51 \cdot 50 \cdot 49 \ldots 3 \cdot 2 \cdot 1$. Even though this number is much smaller than 52^{52} and has "only" 68 digits, we still don't want to write out this number. Furthermore, writing out this 68-digit number would not convey anything about where this number came from and would not give any insight into the problem. It seems we need some new notation, kind of like exponent notation but instead of multiplying by the same number each time, we decrease the number we are multiplying by each time, as in $52 \cdot 51 \cdot 50 \cdot 49 \ldots 3 \cdot 2 \cdot 1$.

DEFINITION

The symbol "!", read aloud as "**factorial**," means the product of all whole numbers starting from the number indicated down to 1. For example, 3!, read aloud as "three factorial," means $3 \cdot 2 \cdot 1$, so $3! = 6$.

So, rather than answering the question in Example 1.2d as $52 \cdot 51 \cdot 50 \cdot 49 \ldots 3 \cdot 2 \cdot 1$, we can say the answer is **52!**. In addition to making the answer easier to write, this notation also helps give some insight into the solution to the problem: take all the whole numbers starting from 52 down to 1 and multiply them together.

We can see now that the difference between selecting with replacement and without replacement becomes larger as we select more and more objects. Although 52^{52} and 52! are both very large numbers, $\frac{52!}{52^{52}} \approx 4.72579 \times 10^{-22}$, which means the monstrous list of all 52! arrangements when repetition is not allowed takes up a miniscule 0.0000000000000000000472% of the ridiculously larger gigantic list of all 52^{52} arrangements if repetition is allowed.

For our purposes, we will consider factorial notation as defined for positive whole numbers only. (It turns out, however, that mathematicians have discovered/invented a function that allows them to perform strange factorial calculations, like $\frac{1}{2}!$. Weird, no?) Also, we will define $0! = 1$ for reasons that may be clear shortly.

Factorial notation is certainly helpful if we are selecting every member of a group, like all 52 cards in a deck. But, what if we are not selecting every object from the group? Can we still use factorial notation to help write the answer?

Example 1.2e: A computer program generates access codes consisting of five letters A–Z, but each letter can be used only once per access code. How many different codes are possible?

Solution: There are 26 choices for the first letter, but then only 25 choices for the second letter, and 24 for the third, 23 for the fourth, and 22 for the fifth. Certainly we can write $26 \cdot 25 \cdot 24 \cdot 23 \cdot 22$, but this isn't particularly satisfying and would get worse if the codes contained more letters (say, for example, 20—Bleh!). The numerical value of the answer (7,893,600) doesn't give us any insight into the problem. How can we use factorial notation to write this number?

26! would mean $26 \cdot 25 \cdot 24 \cdots 3 \cdot 2 \cdot 1$, but we don't want all of these numbers multiplied together; we just want them down to $26 \cdot 25 \cdot 24 \cdot 23 \cdot 22$. All of the numbers from 21 down to 1 need to be canceled out. To cancel out a multiplication, we use division. Therefore, we write

$$26\ 25\ 24\ 23\ 22 = \frac{26 \cdot 25 \cdot 24 \cdot 23 \cdot 22 \cdot 21 \cdot 20 \ldots 3 \cdot 2 \cdot 1}{21 \cdot 20 \ldots 3 \cdot 2 \cdot 1} = \frac{26!}{21!}.$$

Not only does this use of factorial notation make the answer easier to write, it also gives us some insight into the answer: 26 objects are in our group, and we want to select five of them without replacement (and the order of selection matters), so 21 objects are being left out.

Does this idea extend to other problems?

Example 1.2f: 15-character tracking numbers are generated using letters A–Z and numbers 1–9 (0 is omitted to avoid confusion with the letter O). If each character can only be used once per tracking number, how many different tracking numbers are possible?

Solution: One option is to write out $35 \cdot 34 \cdot 33 \cdot 32 \cdots$. But, what is the last number we would need, anyway? If 15 characters are being used, that means 20 are being left out, which means there will be 21 choices for the last character. So our answer is $35 \cdot 34 \cdot 33 \cdots 23 \cdot 22 \cdot 21$. Is there a better way to write this? More efficiently, we say the answer is $\frac{35!}{20!}$.

This type of problem and solution structure occurs frequently enough that mathematicians have given it a name: **permutations**.

DEFINITION

A **permutation** is an arrangement of objects from a group where no object can be used more than once and the order of selection matters.

PERMUTATIONS FORMULA

The total number of permutations of k objects from a group containing n objects is given by the formula $\frac{n!}{(n-k)!}$.

This formula should make good sense if we abstractly consider a problem like the tracking number problem. Given n objects, if all of the objects are arranged, this can happen in $n!$ ways. If k of the objects are being arranged, then the first k numbers in the list n, $n-1$, $n-2$, etc., need to be multiplied together. This leaves $n-k$ numbers at the end of the list to be canceled out (since $k+n-k=n$), ending with 1. Therefore, $(n-k)!$ needs to be canceled out of $n!$, and hence $\frac{n!}{(n-k)!}$.

$$\underbrace{n\cdot(n-1)\cdot(n-2)\cdots(n-k+2)\cdot(n-k+1)}_{k\ terms}\cdot\underbrace{(n-k)\cdot(n-k-1)\cdots 3\cdot 2\cdot 1}_{n-k\ terms}$$

Example 1.2g: An art director for a museum is selecting paintings to be displayed along a stretch of hallway. There is space for 5 paintings to be displayed, and the art director has 12 different paintings from which to choose. How many different ways can the art director display 5 paintings down the hallway?

Solution: This is an example of permutations since no painting can be chosen more than once and because the order of selection matters. (Being the first painting down the hallway is different from being the third painting down the hallway.) Therefore, there are $\frac{12!}{7!}$ = **95,040** different displays possible.

But what if we are selecting objects and the order of selection *doesn't* matter; for example, what if we are being dealt three cards out of a standard deck. Typically when we play card games, the order in which we are dealt cards doesn't matter, only which cards are in our hand after the deal. Another classic example is selecting people to be members of a committee: the order of the people selected (usually) doesn't matter, just who has been selected at the end. How can we count arrangements in these types of situations?

1.3 COMBINATIONS

EXAMPLE 1.3A: Three cards are dealt from a standard deck of 52 cards without replacement. How many different sets of three cards are possible? (The order in which the cards are dealt does not matter.) A standard deck of cards is divided into four suits (hearts, diamonds, spades, and clubs) each of which contains 13 cards (numbers 2 through 10, jack, queen, king, and ace).

SOLUTION: We already know there are $\frac{52!}{49!} = 132{,}600$ different arrangements of three cards if the order does matter, and we have been imagining the list of all these possible arrangements. Let's consider an entry in this list, like 2♣, 3♠, 4♦. Since we created this list of arrangements with the understanding that order mattered, these same three cards will appear in other entries in the list, just in different orders. How many entries will this be? Since we have three different cards, it should be $3 \cdot 2 \cdot 1 = 6$ different times. Listing them out confirms this:

2♣, 3♠, 4♦; 2♣, 4♦, 3♠; 3♠, 2♣, 4♦; 3♠, 4♦, 2♣; 4♦, 2♣, 3♠; 4♦, 3♠, 2♣

So, this particular set of three cards was counted six times when order mattered, but should now only be counted one time since order doesn't matter. The tricky part of this argument is convincing ourselves that this is true for *every* set of three cards. There should be nothing special about 2♣, 3♠, 4♦, and we could repeat the same listing of six arrangements for any set of three cards. (Convince yourself this is true.) The list of all 132,600 arrangements when order matters has therefore overcounted by a factor of 6 when order doesn't matter, so our final answer is 132,600 / 6 = **22,100** different possible sets of three cards.

The key idea for solving this problem was that the list of all possible arrangements when order matters (permutations) over-counted the list of all possible arrangements when order doesn't matter (we don't have a name for this yet). In this case, the over-counting factor was 6, because we were selecting three objects, and those three objects could be arranged in $3 \cdot 2 \cdot 1 = 6$ ways. But wait! $3 \cdot 2 \cdot 1 = 3!$, so we could have written this entire calculation as $\frac{52!}{49!3!}$.

This seems like a nice formula, and each portion of the calculation seems to make sense: we have 52 objects, we are selecting 3 of them and leaving out 49 of them, and the order of selection doesn't matter, so we need to take care of the over-counting factor. Will this strategy and formula always work on these types of problems?

Example 1.3b: A committee of four teachers needs to be selected from the 16 math teachers in a school to select a new Algebra 1 textbook. How many different committees are possible?

Solution: This is a mathematically similar situation to being dealt cards from a deck because no object (or in this case, person) can be selected more than once, and the order of selection doesn't matter. If our nice formula from the previous example is going to work here, we hope the calculation is $\frac{16!}{12!4!}$. Let's carefully think through this problem and see if our formula turns out to be correct.

If the order of selection mattered (maybe the order of selection determines the role the person will have within the committee, like chair or recorder), then this problem would be a permutation, and we would have $\frac{16!}{12!} = 43{,}680$ different committees. But, the order of selection doesn't matter, so our list of 43,680 permutations is too long. By what factor have we over-counted?

Let's consider one item in our list of 43,680 possible permutations: *Mr. Hanks, Ms. Roberts, Mr. Jones, Ms. Wright*. How many different times does this group of four teachers appear in the list of all possible permutations? Since there are four different teachers, and they will appear in every possible order, these four teachers will appear in the list of all permutations $4 \cdot 3 \cdot 2 \cdot 1 = 24$ times. (If you are unsure of this, write them out!) Furthermore, *any* set of four teachers will appear in the list of all permutations 24 times, so the over-counting factor is 24. Therefore, the calculation to determine the number of possible committees of four teachers is $\frac{16!}{12!4!}$, as predicted, and there are **1,820** different possible committees.

Based on the discussion of these examples, it seems we are ready to name this type of problem and generalize the calculation.

DEFINITION

A **combination** is an arrangement of objects from a group where no object can be used more than once and the order of selection does not matter.

COMBINATIONS FORMULA

The total number of combinations of k objects from a group containing n objects is given by the formula $\frac{n!}{(n-k)!k!}$.

NOTATION

Mathematicians use the notation $\binom{n}{k}$, read aloud as "n choose k", for the number of possible combinations when k objects are selected from n objects.

Example: There are $\binom{16}{4}$ (read aloud as "16 choose 4") ways to select a committee of four from a group of 16 people. $\binom{16}{4} = \frac{16!}{12!4!}$, so there are 1,820 different possible committees.

The similarity of the combinations formula to the permutations formula should not be surprising, as we used the permutations formula as the starting point to build the combinations formula. The important difference between a permutation and a combination is that in a combination, order does not matter. In a permutation, order does matter. This means that the list of all possible permutations over-counts the list of all possible combinations since different arrangements of objects are considered distinct in a permutation, but are all counted the same in a combination. Hopefully, the discussion of the two previous examples has given some informal justification to the fact that the over-counting factor is $k!$. Formally, if we consider a set of k objects, there are $k!$ different ways to arrange these objects without repetition. Therefore, the set of all permutations will over-count the number of all combinations by exactly $k!$.

Although knowing and understanding the formulas for permutations and combinations is very important, computing permutations and combinations by hand is not strictly necessary and indeed should be avoided in certain situations. (For example, $\binom{100}{50}$ should not be calculated by hand.) Almost all calculators have shortcuts or commands for combinations and permutations, and you should learn how to use these commands efficiently.

Combinations are an extremely useful and flexible mathematical concept with a wide array of applications. Throughout the remainder of this *Mathematics Resource Guide*, we will use combinations in a variety of ways. In particular, we will focus on the use of combinations in algebra and probability. In order to give readers some idea of the utility of combinations at this time, we will conclude this section with an example using combinations in a non-routine way.

Example 1.3c: Toni has decided that giving each of the 15 guests exactly the same number of packages of candy is a bit boring and wants to mix it up a bit. Toni would like everyone to get at least two packages of candy, but the remaining packages of candy will be distributed randomly into the gift bags, so that each guest will probably end up with a different number of packages of candy. How many different ways can Toni distribute the 90 packages of candy?

Solution: This problem seems to be a long way from combinations. Toni could distribute the candy in a seemingly endless number of ways. For example, one person could get 62 packages of candy, and everyone else could get 2. Two people could get 32 packages of candy, and everyone else could get 2. As long as the total number of packages adds up to 90, and there are fifteen bags, this is a possible arrangement.

90 packages of candy and 15 bags is quite a bit of candy to distribute. Let's try the problem with smaller numbers and see what happens.

What if Toni only has 10 packages of candy and 5 guests? Let's say each person will receive at least one package of candy. This means there are only 5 packages of candy that can "float" from bag to bag. Let's write out a few possibilities for these 5 packages being separated into the five different gift bags.

1, 1, 1, 1, 1

0, 2, 0, 1, 2

4, 1, 0, 0, 0

0, 0, 1, 3, 1

Although there are clearly fewer arrangements for this problem than with our original problem, the connection to combinations—or a quick way to determine the number of possibilities—still seems unclear.

Let's say Toni has the five different gift bags in a row and has the five packages of candy that are allowed to "float" in hand. Toni will put some number of packages in the first gift bag (possibly 0), move to the second gift bag, put some number of packages in the second gift bag (again, possibly 0), and so on down the line. If p represents placing a package of candy in the gift bag, and n represents moving to the next gift bag, then each distribution of the 5 packages can be represented by a string of 5 p's and 4 n's. For example:

1, 1, 1, 1, 1 can be represented by $pnpnpnpnp$

0, 2, 0, 1, 2 can be represented by $nppnnpnpp$

4, 1, 0, 0, 0 can be represented by $ppppnpnnn$

0, 0, 1, 3, 1 can be represented by $nnpnpppnp$

Since there is a one-to-one correspondence between the arrangements of p's and n's and the distributions of the packages of candy, there is the same number of arrangements as distributions. With 9 spots to be filled, five of them need to be selected for p. (The remaining spots will automatically be filled by n's.) No spot can be picked more than once, and the order of selection doesn't matter. Ah ha! A combination!

There are therefore $\binom{9}{5} = 126$ different arrangements of 5 p's and 4 n's, and this means there are 126 ways to fill 5 different gift bags with 10 packages of candy, assuming each bag must have at least one package.

Can this thought process be extended to our original problem? With 15 gift bags, if each gift bag must have at least two packages of candy, 60 packages of candy are allowed to "float" from bag to bag. If we imagine the 15 gift bags in a row, Toni will put some number of packages of candy in the first gift bag (possibly 0), move to the second gift bag, put some number of packages of candy in the second bag (possibly 0), move to the third bag, and so on. At the end of the process, Toni will have placed 60 packages of candy and moved to the next gift bag 14 times.

We can then imagine a tremendously long string of p's and n's representing each possible placement of the candy as Toni moves from bag to bag. Each string will contain 74 total characters: 60 p's and 14 n's. Therefore, we must select 60 of the 74 places for the p's, and these places are selected without replacement, and the order of selection does not matter. Again we have a combination! Therefore, there are $\binom{74}{60}$ = **87,178,291,200** different ways Toni can distribute the 90 packages of candy to the 15 gift bags, assuming each gift bag contains at least two packages of candy. Whew! It is a good thing we didn't try to list them all out!

This example illustrates the common "lineup" representation and use of combinations to count the total number of possible arrangements in said lineup. As we will see in other sections of this resource guide, combinations are a useful mathematical tool with a wide array of applications. At this point, we hope you have some sense of when a combination might be useful and knowledge of how to properly calculate combinations.

SECTION 1 SUMMARY: OVERVIEW OF PERMUTATIONS AND COMBINATIONS

- **The Multiplication Principle:** When listing out all the possibilities for k items, the total number of entries in this list is given by $n_1 \cdot n_2 \cdot n_3 \cdots n_k$, where n_k is the number of possibilities for the k^{th} item. (For example, n_3 is the number of possibilities for the third item.)
- **Factorial Notation:** The symbol "!", read aloud as "factorial," means the product of all whole numbers starting from the number indicated down to 1. For our purposes, factorial is only defined for positive whole numbers and 0. By definition, 0! = 1.
- A **permutation** is an arrangement of objects from a group where no object can be used more than once, and the order of selection matters.

- **Permutations Formula:** The total number of permutations of k objects from a group containing n objects is given by the formula $\frac{n!}{(n-k)!}$.

- A **combination** is an arrangement of objects from a group where no object can be used more than once, and the order of selection does not matter.

- **Combinations Formula:** The total number of combinations of k objects from a group containing n objects is given by the formula $\frac{n!}{(n-k)!k!}$.

- **Combinations Notation:** Mathematicians use the notation $\binom{n}{k}$, read aloud as "n choose k," for the number of possible combinations when k objects are selected from n objects. Therefore, $\binom{n}{k}=\frac{n!}{(n-k)!k!}$.

SECTION 1 REVIEW PROBLEMS: OVERVIEW OF PERMUTATIONS AND COMBINATIONS

For each of the problems below, identify whether the problem should be solved using the Multiplication Principle, permutations, combinations, or some mixture of these three methods. Explain your reasoning, and then solve the problem.

1. At the beginning of math class every day, Mr. Smith selects students to write up homework problems on the board. These problems can be discussed as a class. There are 26 students in Mr. Smith's math class, and he randomly selects with replacement a student to write up each of the first five problems. How many different ways can students be assigned to the problems?

2. Mr. Smith's students eventually complain that it isn't fair that a student can be selected more than once and can be selected to write up all five problems. Mr. Smith agrees that he will now select the five students each day without replacement. How many different ways can students be assigned to the problems?

3. Mr. Smith's students now comment that a student may be selected for a problem they are not able to do correctly and propose an alternate selection method: a group of five students is selected at random, and then these five students decide amongst themselves who will write up each problem. Mr. Smith eventually agrees. How many different ways can Mr. Smith select a group of five students?

4. A *regular polygon* is a polygon with equal angle measures and equal side lengths. A *diagonal* of a polygon connects two non-adjacent vertices. How many diagonals are there in a regular heptadecagon (17-sided polygon)?

5. In a computer game for children, a picture of scenery (like a mountainside) is divided into 8 regions. There are 6 different choices of color, and any region can be painted any color. How many different ways is it possible to color a given picture?

6. Ernest is traveling to New York City and has created a list of ten different possible sightseeing activities in which he is interested. He will be in New York City for three days, but will only have time for two different activities each day. How many different sightseeing plans can Ernest create? (Assume each day is treated separately, and clearly Ernest will not want to complete each activity more than once.)

7. The programming director of a local television station is setting the schedule for the upcoming Saturday. On Saturdays, this station shows a series of five movies taking up the morning and afternoon programming hours. If the programming director has 45 movies from which to choose, how many different movie schedules are possible for this Saturday?

8. What is the coefficient of x^9 in $(x + 1)^{13}$?

Section 2

Algebra

Unlike permutations and combinations, algebra is a topic that usually receives a great deal of discussion in a typical high school mathematics sequence. As with all branches of mathematics, however, many ideas remain beyond what most students encounter during their high school years. In this section of the resource guide, we hope to broaden the scope of traditional high school algebra and show how some of the different mathematical ideas in and around algebra are related to each other. We assume the reader has some familiarity and fluency with topics traditionally covered in high school algebra courses, such as solving linear equations, basic factoring, solving quadratics, and the quadratic formula. Rather than rehash topics such as these, which are typically discussed in traditional high school math classes, we will focus on some topics that may not receive as much attention but are still critically important in higher mathematics: sequences and series, polynomials, the Binomial Expansion Theorem, compound interest, and Euler's constant.

2.1 SEQUENCES AND SERIES

Mathematics, in part, is the study of patterns and attempts to recognize, extend, and classify these patterns. They may occur geometrically, graphically, or algebraically, but many of the patterns in which mathematicians are interested are numerical ones. From an early stage in our mathematical studies, we are familiar with lists of numbers and attempts to extend these lists.

Example 2.1a: What are the next three numbers in the following list: 2, 4, 6, 8 …

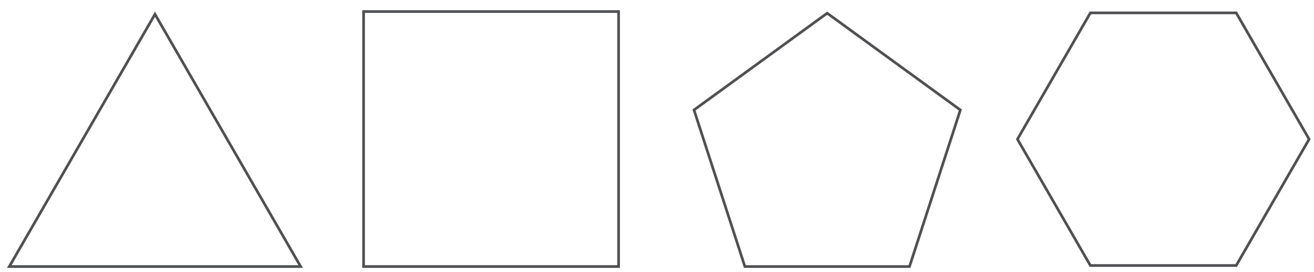
Solution: As this seems to be a list of even numbers, the next three numbers are **10, 12, 14**. Mathematicians call a list of numbers like this a *sequence*.

DEFINITION

A **sequence** is a list of objects presented in a particular order. The objects in the sequence are called the **terms** of the sequence.

Most sequences are made up of numbers, as in Example 2.1a, and most sequences we will encounter in this resource guide will be numerical sequences, but strictly speaking, the terms of a sequence do not need to be numbers. Mathematicians often study sequences of different sorts of objects.

Consider, for example, what object comes next in the following sequence:



Or, for another example, what object comes next in the following sequence: *Monday, Tuesday, Wednesday, Thursday…*?

All of the sequences we will consider are "nice" in that the sequence can be extended in a logical manner. Mathematicians do consider lists that do not have a pattern or formula to be sequences (a sequence of random numbers, for example), but we will restrict ourselves to nice sequences that we can extend logically and predict.

Even within the realm of numerical sequences that can be extended, there are problems assuming that sequences are what they first appear to be. Any numerical sequence can, in theory, be extended in an infinite number of different ways that may or may not agree with the perceived rule or pattern for the sequence. For example, we said earlier that 2, 4, 6, 8, … appears to be a list of even numbers, and so extended it with 10, 12, and 14. However, this sequence does not have to be a list of even numbers, and there are many different ways to extend this sequence. The sequence 2, 4, 6, 8, 10, 58, 252, 734… is a valid mathematical sequence, although it is very different from how we expect a sequence beginning with 2, 4, 6, 8 to continue. In this resource guide, if a sequence seems to follow a simple pattern, we will generally assume it continues this pattern, but strictly speaking this is not necessarily the case.

In order to communicate effectively about sequences, mathematicians have developed some common notation.

NOTATION

The position of a term in a sequence is called the **index** of the term. The terms of a sequence are denoted by a variable (usually a or x) and an index, with the index written as a subscript. Unless otherwise noted, the index begins at 1 and consists of counting numbers.

For example, a generic sequence will commonly be written as $x_1, x_2, x_3\ldots$ or $a_1, a_2, a_3\ldots$ For the sequence of even numbers given earlier, $x_1 = 2$, $x_2 = 4$, $x_3 = 6$, $x_4 = 8$, etc.

Often when we write a sequence in this form, it becomes apparent that there is a relationship between the index and the term of the sequence. In the sequence of even numbers, the term is always twice the index.

Mathematicians will use another variable, usually i or k, to represent the index, and x_i or x_k will be used to represent the i^{th} or k^{th} term in the sequence. When this is done, a formula can be written to describe the relationship between the index and the term. We will use curly brackets {} to denote that an equation represents a sequence of terms, rather than an equation we might try to solve.

For example, the sequence of even numbers can be written as $\{x_i = 2i\}$.

The curly brackets tell us we are talking about a sequence, and the index of this sequence is represented by i. With no other indication about what values i takes, we assume i begins at 1 and counts upward. When $i = 1$, $x_1 = 2 \cdot 1$, so $x_1 = 2$. When $i = 2$, $x_2 = 2 \cdot 2$, so $x_2 = 4$. When $i = 3$, $x_3 = 2 \cdot 3$, so $x_3 = 6$. This pattern continues, so this sequence is the sequence of even numbers.

EXAMPLE 2.1B: What sequence is generated by $\{x_i = i^2\}$?

SOLUTION: When it is not clear what a sequence is, writing out several terms is always a good place to start. When $i = 1$, $x_1 = 1^2$, so $x_1 = 1$. When $i = 2$, $x_2 = 2^2$, so $x_2 = 4$. When $i = 3$, $x_3 = 3^2$, so $x_3 = 9$. When $i = 4$, $x_4 = 4^2$, so $x_4 = 16$. The first four terms of the sequence are 1, 4, 9, 16, and the fifth term will be 25 (verify!), so this is the sequence of square numbers.

EXAMPLE 2.1C: What formula will generate the sequence 3, 7, 11, 15, 19, …?

SOLUTION: For this sequence, $x_1 = 3$, $x_2 = 7$, $x_3 = 11$, and $x_4 = 15$. It seems the index is multiplied by 4 and then one is subtracted to give the term. So, the formula that generates this sequence is $\{x_i = \mathbf{4}\boldsymbol{i} - \mathbf{1}\}$.

Although all of the sequences we have considered so far go on forever, or are **infinite**, many sequences in which we are interested do not go on forever, or are **finite**. A finite sequence only has a certain number of terms, so the index will have a starting value and ending value. We denote these limits on the index by using a superscript and subscript after the second curly bracket. The lower number represents the starting value for the index (sometimes accompanied by "i = "), and the upper number represents the ending value for the index.

EXAMPLE 2.1D: Write out the terms of the finite sequence $\{x_i = \sqrt{i}\}_{i=1}^{9}$.

SOLUTION: Simplifying the square roots as we go, this sequence is 1, $\sqrt{2}$, $\sqrt{3}$, 2, $\sqrt{5}$, $\sqrt{6}$, $\sqrt{7}$, $\sqrt{8}$, 3.

Some sequences have no particular pattern, or clear relationship of one term to the next, and a generating formula must be "guessed" or intuited somehow. Two particular types of sequences, however, have nice relationships from one term to the next, and therefore have nice generating formulas. These types of sequences occur frequently and have nice mathematical characteristics, which suggests they merit their own terminology and careful study.

2.1.1 ARITHMETIC AND GEOMETRIC SEQUENCES

Sequences are particularly nice if there is a relationship that is easy to identify between one term and the next. Let's take a look at an example: What is the apparent relationship between the terms in the sequence 3, 7, 11, 15, 19…? As mentioned previously, just because a sequence appears to have a pattern or relation ship does not necessarily mean that pattern holds in the sequence. If we assume this sequence behaves as it seems to, the relationship is easy to identify: to get from one term to the next, we add 4.

We have been discussing formulas that generate the terms of a sequence, like $\{x_i = 4i - 1\}$ or $\{x_i = i^2\}$. But there is another way to represent a sequence as a formula, and that is to describe how the sequence changes from one term to the next. In this case, the sequence is adding 4 to go from one term to the next. Mathematically, we write $x_i = x_{i-1} + 4$. Since x_i represents the term in the i^{th} place in the sequence, x_{i-1} represents the term in the $(i-1)^{th}$ place in the sequence, or the previous term.

This formula alone, $x_i = x_{i-1} + 4$, is not enough to describe the sequence since it only describes how to get from one term to the next. This relationship could describe any of a whole group of different sequences that add 4 each time, but begin at different values. Therefore, to properly describe a sequence, this type of formula also needs a declaration of the starting value. In the example above, we would declare $x_1 = 3$. This type of formula for a sequence is called a **recursive formula**.

DEFINITION

A **recursive formula** for a sequence is a formula that declares the starting value (or values) for the sequence and how the subsequent terms are made from the previous term (or terms).

EXAMPLE 2.1E: Write a recursive formula for the sequence 3, 7, 11, 15, 19… .

SOLUTION: The recursive formula is $\boldsymbol{x_1 = 3; x_i = x_{i-1} + 4}$.

EXAMPLE 2.1F: Write a recursive formula for the sequence 3, 6, 12, 24, 48 … .

SOLUTION: The first term of this sequence is also 3; therefore, $x_1 = 3$. What is happening as we move from term to term in this sequence? It appears as though each term is twice the previous term, so $x_i = 2 \cdot x_{i-1}$. Therefore, the recursive formula is $\boldsymbol{x_1 = 3; x_i = 2 \cdot x_{i-1}}$.

Recursive formulas are sometimes easier to write than formulas that generate a sequence directly, like $\{x_i = 4i - 1\}$. These types of formulas are called **direct formulas**. However, to find the 100th term of a sequence using a recursive formula requires writing out the first 99 terms of this sequence, a potentially tedious task. To find the 100th term of a sequence using a direct formula, we substitute $i = 100$ into the formula. The advent of computing technology has somewhat lessened this drawback of recursive formulas, as computers can now compute thousands of terms using a recursive formula fairly quickly. Indeed, a great deal of computer programming uses recursive formulas, and some sequences are much easier to write using recursive formulas than direct formulas.

EXAMPLE 2.1G: What are the terms of the sequence given by the following recursive formula: $a_1 = 1$, $a_2 = 1$; $a_i = a_{i-1} + a_{i-2}$?

SOLUTION: Since they were declared, the first two terms of the sequence are $a_1 = 1$ and $a_2 = 1$. What do we make of the recursive declaration in this formula? We already know a_1 and a_2, so letting $i = 3$ should allow us to find a_3. By substitution, $a_3 = a_{3-1} + a_{3-2}$, so $a_3 = a_2 + a_1$. Therefore, $a_3 = 2$. When $i = 4$, $a_4 = a_3 + a_2$, so $a_4 = 3$. When $i = 5$, $a_5 = a_4 + a_3$, so $a_5 = 5$. Thus far our sequence is 1, 1, 2, 3, 5, and we are able to understand what the recursive rule is doing: each term is found by adding the two previous terms together. Writing out more terms of the sequence gives us 1, 1, 2, 3, 5, 8, 13, 21, 34, … . This famous sequence is called the **Fibonacci sequence**. Although somewhat strange at first, the recursive formula for the Fibonacci sequence is fairly straightforward; the direct formula, on the other hand, is difficult to determine and extremely complicated!

Let's return to less complicated sequences. A sequence where a fixed amount is added to move from one term to the next (like 3, 7, 11, 15, 19, …) is called an **arithmetic sequence.**

DEFINITION

An **arithmetic sequence** is a sequence with a constant difference between consecutive terms.

Although we generally think of this constant difference as a positive number, this is not strictly necessary, and sequences with a constant negative difference are also arithmetic. Sequences with a common difference of 0 are also technically considered arithmetic although these sequences aren't very interesting. (Why not?)

EXAMPLE 2.1H: What is the constant difference between the terms of the arithmetic sequence 13, 6, –1, –8, …?

SOLUTION: The constant difference is **–7**.

NOTATION FOR ARITHMETIC SEQUENCES

The constant difference for an arithmetic sequence is usually represented by the variable d. The first term of an arithmetic sequence is usually represented by the variable a.

This notation allows us to develop recursive and direct formulas for arithmetic sequences. We began this discussion with recursive formulas precisely because the recursive formula for an arithmetic sequence is so nice. The first term is a, so $x_1 = a$. To move from one term to the next, we add d, so $x_i = x_{i-1} + d$.

RECURSIVE FORMULA FOR AN ARITHMETIC SEQUENCE

The recursive formula for an arithmetic sequence with an initial value of a and a constant difference of d is $x_1 = a$; $x_i = x_{i-1} + d$.

Using the recursive formula to write out the terms of the generic arithmetic sequence gives us a, $a + d$, $a + 2d$, $a + 3d$, $a + 4d$, … . A direct formula relies on the relationship between the index and the term, so we consider $x_1 = a$, $x_2 = a + d$, $x_3 = a + 2d$, $x_4 = a + 3d$, and $x_5 = a + 4d$. It seems each term is a plus d times one less than the index. This makes sense if we think about what happens as we move along in the sequence. To move from the first term to the fifth term, we will add the difference, d, four times to the starting value, a. Therefore, the fifth term will be $a + 4d$. Generalizing this argument, to move from the first term to the k^{th} term, we will add the difference, d, $k - 1$ times to the starting value, a. Therefore, the k^{th} term will be $a + (k - 1) \cdot d$.

DIRECT FORMULA FOR AN ARITHMETIC SEQUENCE

The direct formula for an arithmetic sequence with an initial value of a and a constant difference of d is $x_k = a + (k - 1) \cdot d$.

It makes sense that the direct formula for an arithmetic sequence would involve a multiplication by the constant difference, as this difference is added repeatedly to move from term to term, and repeated addition can be expressed as multiplication.

EXAMPLE 2.1I: What is the 201st term in the arithmetic sequence 41, 38, 35, 32, …?

SOLUTION: The first term of the sequence is 41, and the common difference is –3. To move from the 1st term to the 201st term in the sequence, the common difference will be added 200 times, so a total of 600 will be subtracted from 41. Therefore, the 201st term is **–559**. Note that using the direct formula is precisely the same logic, just written more formally: $x_{201} = 41 + (201–1) \cdot (–3)$, so $x_{201} =$ **–559**.

EXAMPLE 2.1J: What is the first term of the arithmetic sequence with $x_{54} = 136$ and $x_{77} = 205$?

SOLUTION: The first thing we need to do is to determine the constant difference. From the 54th term to the 77th term, the difference is added 23 times, and the distance between $x_{54} = 136$ and $x_{77} = 205$ is 69. Therefore, the constant difference is 3. From the first term to the 54th term, the difference was added 53 times. This means the sequence increased 159 from the first term to the 54th term. Therefore, the first term is **–23**.

The thought process in this example is a bit of an informal argument. We can use the direct formula several times in a more mathematically formal way to solve the same problem.

Substituting known values into $x_k = a + (k - 1) \cdot d$ gives us two equations: $136 = a + 53 \cdot d$ (when $k = 54$) and $205 = a + 76 \cdot d$ (when $k = 77$). This is then a system of two equations with two unknowns that can be solved using substitution or elimination. For example, subtracting the two equations gives us 69 = 23d, which implies d = 3. Substituting into the first equation then yields 136 = a + 53(3), which solves as a = **–23**.

The other (mathematically) simple thing to do as we move from one term to the next in a sequence is multiply, as in the sequence 3, 6, 12, 24, 48, … given in Example 2.1f. These types of sequences are called **geometric sequences**.

DEFINITION

A **geometric sequence** is a sequence with a constant ratio between consecutive terms.

Although we usually consider multiplication and division as separate operations, we know that division problems can be stated as multiplication problems and vice versa. Therefore, geometric sequences are always phrased as a multiplication from term to term, although sometimes this ratio is a fraction between 0 and 1.

Example 2.1k: What is the ratio for the geometric sequence 36, 12, 4, $\frac{4}{3}$, …?

Solution: To move from term to term in this sequence, we divide by 3, so the ratio is $\mathbf{\frac{1}{3}}$.

Ratios for geometric sequences may also be negative numbers or 1. (What would these sequences look like?). The only number that is not allowed as a ratio for a geometric sequence is 0. (Why?)

NOTATION FOR GEOMETRIC SEQUENCES

The constant ratio for a geometric sequence is usually represented by the variable r. The first term for a geometric sequence is usually represented by the variable a.

As with arithmetic sequences, this notation allows for the development of recursive and direct formulas for geometric sequences. Since the first term of a geometric sequence is a, $x_1 = a$. Since a geometric sequence multiplies by the ratio from term to term, $x_i = r \cdot x_{i-1}$.

RECURSIVE FORMULA FOR A GEOMETRIC SEQUENCE

The recursive formula for a geometric sequence with an initial value of a and a ratio of r is $x_1 = a$; $x_i = r \cdot x_{i-1}$.

We can use this recursive formula to write out several terms and find a direct formula. The terms of the generic geometric sequence are a, $a \cdot r$, $a \cdot r^2$, $a \cdot r^3$, $a \cdot r^4$… . A direct formula finds a relationship between the index and the term, so we consider $x_1 = a$, $x_2 = a \cdot r$, $x_3 = a \cdot r^2$, $x_4 = a \cdot r^3$, and $x_5 = a \cdot r^4$. It seems that each term is a times r raised to a power that is one less than the index. Therefore, the k^{th} term will be a times r raised to the $k - 1$ power, and $x_k = a \cdot r^{k-1}$.

DIRECT FORMULA FOR A GEOMETRIC SEQUENCE

The direct formula for a geometric sequence with an initial value of a and a ratio of r is $x_k = a \cdot r^{k-1}$.

It makes sense that this formula contains a power of r, as a geometric sequence involves repeated multiplication by r to move from one term to the next. Since repeated multiplication can be represented as an exponent, we anticipate the direct formula would have a power of r. This power should be $k - 1$, since we need one power of r for every term except for the first term.

EXAMPLE 2.1L: What is the 20th term in the geometric sequence $\frac{1}{20}, \frac{1}{10}, \frac{1}{5}, \frac{2}{5}, \ldots$?

SOLUTION: As the first term is $\frac{1}{20}$ and the ratio is 2, the 20th term is given by $x_{20} = \frac{1}{20} \cdot 2^{20-1}$, so the 20th term is $\mathbf{\frac{131072}{5}}$.

EXAMPLE 2.1M: What is the 4^{th} term in a geometric sequence if the 2^{nd} term is 2 and the 6^{th} term is 8?

SOLUTION: Since a geometric sequence multiplies by a constant ratio to move from term to term, this multiplication occurs 4 times to move from the 2^{nd} term to the 6^{th} term. This means that $r^4 = 4$, since the 2^{nd} term times 4 is the 6^{th} term. So, $r^2 = 2$ and $r = \sqrt{2}$. We will multiply by the constant ratio twice to move from the 2^{nd} term to the 4^{th} term, so the 2^{nd} term will be multiplied by $r^2 = 2$, and therefore the 4^{th} term in this sequence is **4**. (The first six terms of the sequence are $\sqrt{2}, 2, 2\sqrt{2}, 4, 4\sqrt{2}, 8$.)

This is again a bit of an informal argument, and it can be made more formal by using the direct formula for a geometric sequence. As $x_2 = 2$, $2 = a \cdot r$, and as $x_6 = 8$, $8 = a \cdot r^5$. Dividing these two equations gives us $r^4 = 4$, and so $r = \sqrt{2}$. We are looking for x_4, which is $a \cdot r^3$. Rather than solving for a, we use $2 = a \cdot r$ to build $a \cdot r^3$ by multiplying both sides by r^2. This gives us $2 \cdot r^2 = a \cdot r^3$. As $r = \sqrt{2}$, $r^2 = 2$, so $4 = a \cdot r^3$, and $x_4 = 4$.

Arithmetic and geometric sequences are mathematically important due to their simplicity and ease of use. They also behave nicely when summing the terms of a sequence, which we will investigate shortly. Arithmetic and geometric sequences are important in modeling physical phenomena and in other applications. Although most models are neither precisely arithmetic (linear) nor geometric (exponential), many growth or decay situations can be modeled extremely well using these simple equations. Some of the models encountered in typical high school mathematics are perhaps a bit simplistic, but these models are extremely powerful and far reaching. For example, the idea of the recursive step for the geometric sequence, $x_i = r \cdot x_{i-1}$, forms a large part of the study of differential equations in higher mathematics.

2.1.2 ARITHMETIC AND GEOMETRIC SERIES

In addition to being mathematically simple and having nice recursive and direct formulas, arithmetic and geometric sequences have another important mathematical property: it is possible to find a nice formula for the *sum* of an arithmetic (or geometric) sequence. This sum is called a **series**.

DEFINITION

A **series** is the sum of the terms in a sequence.

Although it is always possible to find a finite series by brute computational force (determining all the terms in the sequence and then summing), this is not a very mathematical way to proceed, and mathematicians try to avoid doing such things whenever possible. The fact that there is a nice way to find the sum of arithmetic and geometric series further highlights their mathematical importance. We will begin by looking at an extremely simple, but very important, arithmetic series.

EXAMPLE 2.1N: What is 1 + 2 + 3 + … + 97 + 98 + 99 + 100?

SOLUTION: The first thing we notice is that this series, like most we will study, is finite; that is, it has a certain number of terms (in this case, 100). Many sequences we consider are infinite (go on forever), but most infinite series *diverge*, or add up to infinity. For example, any infinite arithmetic series adds up to infinity (like 3 + 7 + 11 + 15 + …) or negative infinity (like 41 + 38 + 35 + 32 + …). However, here we have a finite number of terms, so these 100 numbers certainly add up to some number.

We would like a clever way to add up all 100 of these numbers, rather than just 1 + 2 = 3, 3 + 3 = 6, 6 + 4 = 10, etc. This method is rather boring, and with 99 calculations to perform, we are likely to make a mistake. Rather than adding 1 + 2, why not add 1 + 100? 2 can then be paired with 99, 3 with 98, and so on. All of these pairs have the same sum, 101, and there are precisely 50 pairs (since there are 100 numbers and we are pairing them up). Therefore, this sum is equal to 101 · 50 = **5,050**.

Mathematical legend says the great mathematician Carl Gauss (1777–1855) was in his second-grade class when the teacher gave this problem in the hopes of having thirty minutes of peace and quiet as all his pupils painstakingly carried out all ninety-nine sums. Gauss thought about the problem for a bit and then wrote down nothing but the correct answer, having paired the numbers and multiplied in his head. Whether or not this is true, it exemplifies the idea of *work smarter, not harder.*

Will this pairing idea work on any arithmetic series? If so, it seems we have discovered a powerful tool for finding arithmetic series.

Example 2.1o: Find the arithmetic series 3 + 7 + 11 + 15 + … + 399 + 403 + 407.

Solution: The pairing idea certainly starts off nicely: 3 + 407 = 410, 7 + 403 = 410, 11 + 399 = 410, and so on. But how many pairs of 410 are created by the terms in this series? We need to determine how many terms are in the sequence 3, 7, 11, 15 … 399, 403, 407. The sequence has a common difference of 4, and a total of 404 is added from the first term to get to the last term. This means 4 is added 101 times to get from the first term to the last term, which means there are 102 terms in the sequence (as 4 was not added to get the first term, 3). Therefore there are 51 pairs of 410 in this series, and the series totals to $410 \cdot 51 =$ **20,910**.

Does this series have to be arithmetic for this idea to work? Does this work on other types of series?

Example 2.1p: Determine the sum 1 + 4 + 9 + 16 + 25 + 36 + 49 + 64.

Solution: Although we could just add up all these terms, we are interested in whether or not the pairing strategy works for non-arithmetic series, so the numerical answer here is not the point of the problem. This series is certainly non-arithmetic, since the difference in the first two terms is 3, but the difference between the 2nd and 3rd term is 5, and then 7, etc. Will the pairing strategy work? 1 + 64 = 65, but 4 + 49 = 53. Uh oh. 9 + 36 = 45, and 16 + 25 = 41. None of these pairs have the same sum, so our pairing strategy will not work on this series.

Why does the pairing strategy work on arithmetic series? Since the difference between consecutive terms is always a constant, d, the pair sum increases by d as the index increases by one (we move one term up the sequence). Meanwhile, the pair sum decreases by d as the index decreases by one (we move one term down in the sequence). Therefore, the paired sum remains constant throughout, and our pairing strategy works.

Can we write a formula that describes this pairing strategy for arithmetic series? If we have an even number of terms, the formula should work out nicely. With an even number of terms, we know each term in our series has a pair, and it is easy to determine the number of paired sums. Therefore, we will state our formula at first only for arithmetic series with an even number of terms.

STRATEGY AND FORMULA FOR AN ARITHMETIC SERIES WITH AN EVEN NUMBER OF TERMS

To find an arithmetic series with an even number of terms, we create paired sums equal to the first term plus the last term. If there are k terms, there will be $\frac{k}{2}$ pairs, so the arithmetic series will be $(x_1 + x_k) \cdot \frac{k}{2}$.

As $x_1 = a$ and $x_k = a + (k-1) \cdot d$, this formula can also be written as $[2a + (k-1) \cdot d] \cdot \frac{k}{2}$, although this lacks some of the mathematical aesthetic of the first formula, as it involves the same number of variables and masks the idea that generated the formula.

What if the arithmetic series has an odd number of terms? In such cases, our pairing strategy will work for the most part, but one term in the series will be left out and not have a pair. How can we deal with this problem? It seems there are three main ways to deal with this problem. One idea is to add an extra term to the series (giving it an even number of terms) and then subtract the additional term off to return to the original series. Another idea is to add up all of the terms except for the last term, so that we have an even number of terms, and then add on the term we left off. The last way is to pair up the terms as usual and then try to figure out the middle term that is left over and has no pair. We encourage you to try the second and third strategy; we will focus on the first.

Let's try our idea with a specific arithmetic sequence to see how it works before we try it in general.

EXAMPLE 2.1Q: What is the arithmetic series $41 + 38 + 35 + \ldots + (-1) + (-4) + (-7)$?

SOLUTION: The constant difference for this series is -3, and there are 17 terms. (Is this correct? How do we know this?) Although our pairing strategy will work, it would be much nicer if there was an an additional term in this series, so there would be 18 terms and 9 pairs. Therefore, we will add –10 to the end of the series, find the sum, and then add 10 to remove this extra term we included. 41 + 38 + 35 + … + (–1) + (–4) + (–7) + (–10) has a paired sum of 31, and there will be 9 of these pairs, so the series is 31 · 9 = 279. But this includes the –10 we added, so our original arithmetic series is **289**.

What if we hadn't worried about whether this series had an even or odd number of terms, and just used $(x_1 + x_k) \cdot \frac{k}{2}$? 41 + (–7) = 34? There would be 8.5 "pairs" since 17/2 = 8.5. And 34 · (8.5) = 289! Amazing! The formula we wrote for an even number of terms seems to work for an odd number of terms as well! Is this just coincidence? Let's try our strategy of including an extra term in general and see what happens.

EXAMPLE 2.1R: What is the arithmetic series with an odd number of terms?

SOLUTION: Since we're doing this problem in general, we'll start with $x_1 = a$, $x_2 = a + d$, and so on up to $x_k = a + (k - 1) \cdot d$, where k is odd. Adding one additional term at the end of this sequence will give us an even number of terms, so we'll include $x_{k+1} = a + k \cdot d$. Now we consider the arithmetic series $a + (a + d) + (a + 2d) + \ldots + [a + (k - 1) \cdot d] + (a + k \cdot d)$. This series has $k + 1$ terms, which is even as k is odd, so this series will equal $(x_1 + x_{k+1}) \cdot \frac{k+1}{2}$, or $(a + a + k \cdot d) \cdot \frac{k+1}{2}$. However, this is not the series in which we are interested, and we need to remove the term we added on. Therefore, our original series will equal $(a + a + k \cdot d) \cdot \frac{k+1}{2} - (a + k \cdot d)$. At this point some algebraic manipulation seems necessary:

$\frac{(2a + k \cdot d) \cdot (k + 1)}{2} - a - k \cdot d$ (we write the first term as a single fraction and distribute the negative across the second term)

$\frac{2ak + 2a + k^2d + k \cdot d}{2} - a - k \cdot d$ (expanding the numerator of the fraction)

$\frac{2ak + 2a + k^2d + k \cdot d}{2} - \frac{2a}{2} - \frac{2k \cdot d}{2}$ (creating a common denominator, so the terms can be combined into one fraction)

$\frac{2ak + k^2d - k \cdot d}{2}$ (subtracting the fractions)

$\frac{(2a + kd - d) \cdot k}{2}$ (factoring out the common factor of k from each term)

$(2a + kd - d) \cdot \frac{k}{2}$ (rewriting multiplication)

$[2a + (k - 1) \cdot d] \cdot \frac{k}{2}$ (factoring out d from the second and third terms inside parentheses)

$[a + a + (k - 1) \cdot d] \cdot \frac{k}{2}$ (splitting apart $2a$ into $a + a$)

$[x_1 + x_k] \cdot \frac{k}{2}$ (substituting $x_1 = a$ and $x_k = a + (k - 1) \cdot d$)

But this is exactly the same formula we had when the series had an even number of terms! Amazing! So, the same formula works whether our arithmetic series has an even number of terms or an odd number of terms.

FORMULA FOR AN ARITHMETIC SERIES

An arithmetic series with k terms is equal to $(x_1 + x_k) \cdot \frac{k}{2}$.

EXAMPLE 2.1S: An arithmetic series equals 624. The first term of this series is 3, and the second term is 5. What is the last term in this series?

SOLUTION: We know that $x_1 = 3$ and $d = 2$. The fact that the series equals 624 means that $(x_1 + x_k) \cdot \frac{k}{2} = 624$, but we only know x_1 and are trying to find x_k. As $x_k = a + (k - 1) \cdot d$ in general, for this problem $x_k = 3 + 2 \cdot (k - 1)$, or $x_k = 1 + 2k$. Substituting in now yields $(3 + 1 + 2k) \cdot \frac{k}{2} = 624$, or $(4 + 2k) \cdot k = 1248$. This is a quadratic in terms of k, so rearranging gives us $2k^2 + 4k - 1248 = 0$ or $k^2 + 2k - 624 = 0$. A quick check of the factors of 624 reveals 26 and 24, so $(k + 26) \cdot (k - 24) = 0$, and $k = -26$ or $k = 24$. Certainly our arithmetic series should have a positive number of terms, so there are 24 terms in our series. This means the last term $x_k = 1 + 2k$ is **49**.

What about geometric series? As we just saw, the pairing trick only works for arithmetic series. Can something be done to make geometric series easy to find?

EXAMPLE 2.1T: Find the geometric series 3 + 6 + 12 + 24 + 48 + 96 + 192 + 384.

SOLUTION: Although this series does not have very many terms, and we could just add them all up, this would not lend itself to a general method. Let's try something else. To find an arithmetic series with an odd number of terms, we added on the next term, which changed the series into something nicer. Then we went back from that series to the original series in which we were interested. Is adding a term to this geometric series helpful?

We are interested in 3 + 6 + 12 + 24 + 48 + 96 + 192 + 384, so let's call this number S. Consider 3 + 6 + 12 + 24 + 48 + 96 + 192 + 384 + 768. This is clearly S + 768, so: 3 + 6 + 12 + 24 + 48 + 96 + 192 + 384 + 768 = S + 768. Subtracting 3 from both sides yields: 6 + 12 + 24 + 48 + 96 + 192 + 384 + 768 = S + 768 – 3.

Now every term on the left-hand side shares a factor of 2, so factoring this out gives us: 2(3 + 6 + 12 + 24 + 48 + 96 + 192 + 384) = S + 768 – 3. But wait! 3 + 6 + 12 + 24 + 48 + 96 + 192 + 384 = S, so: $2S$ = S + 768 – 3.

This equation can easily be solved for S, and so this geometric series is **765**. Amazing! But will this trick always work?

EXAMPLE 2.1U: Find the geometric series $36 + 12 + 4 + \frac{4}{3} + \frac{4}{9} + \frac{4}{27} + \frac{4}{81} + \frac{4}{243}$.

SOLUTION: Let $36 + 12 + 4 + \frac{4}{3} + \frac{4}{9} + \frac{4}{27} + \frac{4}{81} + \frac{4}{243} = S$. The ratio for this geometric series is $\frac{1}{3}$, so the next term would be $\frac{4}{729}$. Let's add this term to our series and then try to rewrite both sides including S.

$$36 + 12 + 4 + \frac{4}{3} + \frac{4}{9} + \frac{4}{27} + \frac{4}{81} + \frac{4}{243} + \frac{4}{729} = S + \frac{4}{729}$$

$$12 + 4 + \frac{4}{3} + \frac{4}{9} + \frac{4}{27} + \frac{4}{81} + \frac{4}{243} + \frac{4}{729} = S + \frac{4}{729} - 36$$

Now the tricky part. We need to rewrite the left-hand side as the original series. In order to turn 12 into 36 and 4 into 12, we need to factor out $\frac{1}{3}$ and write this as:

$$\frac{1}{3}\left(36 + 12 + 4 + \frac{4}{3} + \frac{4}{9} + \frac{4}{27} + \frac{4}{81} + \frac{4}{243}\right) = S + \frac{4}{729} - 36$$

$$\frac{1}{3}S = S + \frac{4}{729} - 36$$

Solving for S gives us: $\frac{26240}{729} = \frac{2}{3}S$, so $S = \mathbf{\frac{13120}{243}}$.

Let's try this strategy in general on a generic geometric series and see if we can develop a formula.

EXAMPLE 2.1V: What is a geometric series with k terms and a ratio of r?

SOLUTION: Consider the generic geometric series $a + a \cdot r + a\,r^2 + \ldots + a \cdot r^{k-2} + a \cdot r^{k-1}$. Let this equal S. The next term will be $a \cdot r^k$, so: $a + a \cdot r + a \cdot r^2 + \ldots + a \cdot r^{k-2} + a \cdot r^{k-1} + a \cdot r^k = S + a \cdot r^k$.

Moving the a to the other side of the equation gives us:

$$a \cdot r + a \cdot r^2 + \ldots + a \cdot r^{k-2} + a \cdot r^{k-1} + a \cdot r^k = S + a \cdot r^k - a.$$

Every term on the left-hand side now has a factor of r, so factoring this out yields:

$$r \cdot (a + a \cdot r + \ldots + a \cdot r^{k-3} + a \cdot r^{k-2} + a \cdot r^{k-1}) = S + a \cdot r^k - a$$

But $a + a \cdot r + a \cdot r^2 + \ldots + a \cdot r^{k-2} + a \cdot r^{k-1}$ is S, so:

$$r \cdot S = S + a \cdot r^k - a$$

Now we solve for S:

$$r \cdot S - S = a \cdot r^k - a$$

$$S \cdot (r - 1) = a \cdot r^k - a$$

$$S = \boldsymbol{\frac{a \cdot r^k - a}{r - 1}}$$

FORMULA FOR A FINITE GEOMETRIC SERIES

A geometric series with first term a, ratio r, and with k terms is equal to $S = \frac{a \cdot r^k - a}{r - 1}$. Alternatively, we can write this formula by referencing the terms in the sequence: $S = \frac{x_{k+1} - x_1}{r - 1}$, where x_k is the last term in the series and x_{k+1} is the next term in the sequence (but is not included in the series).

This formula does not hold if the ratio is equal to 1 because the method used does not work. (Why not? What is the sum if $r = 1$?)

EXAMPLE 2.1W: What is the value of the geometric series $1 + \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{512} + \frac{1}{1024}$?

SOLUTION: Although we could determine the number of terms in this series and use $S = \frac{a \cdot r^k - a}{r - 1}$, since we know the last term in this series, we will use $S = \frac{x_{k+1} - x_1}{r - 1}$. The ratio is $\frac{1}{2}$, and as $x_k = \frac{1}{1024}$, $x_{k+1} = \frac{1}{2048}$. Therefore, the series equals $\dfrac{\frac{1}{2048} - 1}{\frac{1}{2} - 1} = \mathbf{\frac{2047}{1024}}$.

This example brings up an interesting question: what happens if our geometric series has an infinite number of terms? Clearly if an arithmetic series has an infinite number of terms, the series will equal positive infinity (if $d > 0$) or negative infinity (if $d < 0$). If the absolute value of a geometric series increases (if $|r| > 1$), then the absolute value of the series will equal positive infinity. But, what happens if the terms in a geometric series get closer to 0 (if $|r| < 0$), as in this example?

EXAMPLE 2.1X: What is the value of the infinite geometric series $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots$?

SOLUTION: Since this is an infinite geometric series with a ratio of $\frac{1}{2}$, the terms of the sequence are approaching 0. This means the "last term" in this series is 0. (Technically, there is no last term since the series goes on forever, but the terms get close enough to 0 that we can't tell the difference. This will seem a lot less like hand waving if the reader has studied limits.)

Using $S = \frac{x_{k+1} - x_1}{r - 1}$ $x_{k+1} = 0$ and $x_1 = 1$, so $S = \frac{0 - 1}{\frac{1}{2} - 1}$, and $\mathbf{S = 2}$.

In fact, as long as $x_k \approx 0$, $x_{k+1} \approx 0$ as well, and the infinite geometric series will have a finite sum. In this case, $S = \frac{x_{k+1} - x_1}{r - 1}$ simplifies to $S = \frac{-x_1}{r - 1}$ which we can rewrite as $S = \frac{x_1}{1 - r}$.

FORMULA FOR INFINITE GEOMETRIC SERIES

An infinite geometric series equals a finite number if $x_k \approx 0$ for sufficiently large values of k. In this case, $S = \frac{a}{1 - r}$.

EXAMPLE 2.1Y: Determine the value of the infinite geometric series $2 - \frac{4}{3} + \frac{8}{9} - \frac{16}{27} + \frac{32}{81} - \ldots$

SOLUTION: The initial value of this geometric sequence is 2, and $r = -\frac{2}{3}$. Since $x_k \approx 0$ as k increases, this infinite geometric series is a finite value given by the formula $S = \frac{a}{1 - r}$. Therefore, $S = \frac{2}{1 - \left(-\frac{2}{3}\right)}$, and $S = \mathbf{\frac{6}{5}}$.

Arithmetic and geometric series are not the only series that have formulas for determining their values, but they are the two types of series that often occur in applied problems. We will see a few of these applications throughout the *Mathematics Resource Guide*.

2.1.3 SIGMA NOTATION

Writing out a series term by term can be very tedious, especially if there are many terms in the series. Even if we write out the series using an ellipsis, as in 3 + 7 + 11 + 15 + … + 399 + 403 + 407, we still don't know how many terms are in the series, a piece of information we usually would like to have.

Recall our notation for sequences: $\{x_i = 4i - 1\}$ means start with $i = 1$ and then generate terms by increasing the value of the index. Therefore, $\{x_i = 4i - 1\}$ represents the sequence 3, 7, 11, 15, … . Note that this sequence goes on forever, whereas the series 3 + 7 + 11 + 15 + … + 399 + 403 + 407 does not. We determined there were 102 terms in this sequence, so $\{x_i = 4i - 1\}_{i=1}^{102}$ generates the sequence that corresponds to this series. But how do we denote that we are interested in the series as opposed to the sequence?

NOTATION

Mathematicians use sigma notation, $\sum$, to denote a series. $\sum_{i=a}^{b} f(i)$ means the series that corresponds to the sequence generated by the formula $f(i)$ where the index begins at a and ends at b.

For example, to express the series 3 + 7 + 11 + 15 + … + 399 + 403 + 407 in sigma notation, we write $\sum_{i=1}^{102} 4i - 1$. This notation says to create the sequence given by $\{x_i = 4i - 1\}_{i=1}^{102}$, and then add all the terms together to create the series. Sometimes parentheses are used around the formula that generates the terms to avoid potential confusion, as in $\sum_{i=1}^{102} (4i - 1)$.

EXAMPLE 2.1Z: Write the following series in sigma notation: 1 + 4 + 9 + 16 + … + 196 + 225 + 256.

SOLUTION: The terms of this series are consecutive squares, so $x_i = i^2$. The first term is 1^2, and the last term is 16^2. We write $\sum_{i=1}^{16} i^2$

EXAMPLE 2.1AA: Determine the value of $\sum_{k=1}^{10} \frac{1}{k}$.

SOLUTION: This series has 10 terms, since k is going from 1 to 10. The sequence is $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \frac{1}{9}, \frac{1}{10}$, so the corresponding series is the sum of these terms, which equals $\mathbf{\frac{7381}{2520}}$.

Sigma notation also gives us a potentially nicer way to write the formulas for arithmetic and geometric series.

FORMULA FOR ARITHMETIC SERIES [SIGMA FORM]

$$\sum_{i=1}^{k} (a + (i - 1) \cdot d) = [2a + (k - 1) \cdot d] \cdot \frac{k}{2}$$

FORMULA FOR FINITE GEOMETRIC SERIES [SIGMA FORM]

$$\sum_{i=1}^{k} (a \cdot r^{i-1}) = \frac{a \cdot r^k - a}{r - 1}$$

FORMULA FOR INFINITE GEOMETRIC SERIES [SIGMA FORM]

If $a \cdot r^i \approx 0$ for sufficiently large i, $\sum_{i=1}^{\infty} (a \cdot r^{i-1}) = \frac{a}{1 - r}$.

Don't be confused by the use of i and k in these equations. Here i is being used as the index, and k is being used as the last value of the index in the series. Since these formulas represent generic series, the number of terms is unknown and must be represented by a variable that is different from the variable representing the index. Any time there is some confusion about a series, especially in sigma notation, writing out a few terms is always a good idea to get a feel for what is going on. For example, the arithmetic series formula in sigma form may be particularly troublesome. What exactly is going on with $\sum_{i=1}^{k} (a + (i - 1) \cdot d)$? Let's write out a few terms.

When $i = 1$, $a + (i - 1) \cdot d = a$, so $x_1 = a$.

When $i = 2$, $a + (i - 1) \cdot d = a + d$, so $x_2 = a + d$.

When $i = 3$, $a + (i - 1) \cdot d = a + 2d$, so $x_3 = a + 2d$.

When $i = k$, $a + (i - 1) \cdot d = a + (k - 1) \cdot d$, so $x_k = a + (k - 1) \cdot d$.

So this series is $a + (a + d) + (a + 2d) + (a + 3d) + \ldots + (a + (k - 1) \cdot d)$, the generic arithmetic series with k terms.

Sigma notation is an extremely powerful tool that mathematicians use to write out complicated and long sums in nice shorthand. Mastering sigma notation takes a lot of time and practice. If you remember what sigma notation means, you can always write out enough terms to get a feel for what is going on. Now that we have some idea of how to use sigma notation, we will try to utilize sigma notation in our next topic: polynomials.

2.2 POLYNOMIALS

Polynomials are a topic most students encounter in high school mathematics, but perhaps only briefly. Most polynomials encountered in high school math courses are either of degree 1 (linear) or degree 2 (quadratic). Usually degree 3 (cubics) polynomials are introduced, but higher degrees are not usually discussed much, if at all.

Polynomials are also usually studied in the context of functions rather than as entities unto themselves. This means that when we are presented with a polynomial such as $2x^2 + 5x - 3$, we typically think of this as $f(x) = 2x^2 + 5x - 3$ and concern ourselves with the value of the function for different values of x, the graph of the function, the roots of the function, how the factored form of the function relates to the roots, and other questions about the properties of the function. Rather than considering questions such as these, here we will focus on the polynomial itself.

A polynomial is an algebraic object consisting of terms. Each term is made up of a variable, usually x, raised to a different non-negative power and a coefficient, possibly 1 or 0. By convention, the polynomial is written with the powers of x in order, either from highest to lowest or lowest to highest, whichever is more convenient. The highest power of x with a non-zero coefficient is called the **degree** of the polynomial.

The following are examples of polynomials and their degrees:

$2x^2 + 5x - 3$; degree 2

$7x^3 + 7x^2 + 12$; degree 3

$-3x^5 + 4x^4 + 5x^3 - 6x$; degree 5

$1 + x + 2x^2 + 3x^3 + 5x^4 + 8x^5$; degree 5

$1 + \frac{1}{2}x + \frac{1}{4}x^2 + \frac{1}{8}x^3 + \frac{1}{16}x^4 + \frac{1}{32}x^5 + \frac{1}{64}x^6 + \frac{1}{128}x^7$; degree 7

In general, a polynomial looks sort of like a series, but with powers of x attached to each term. For the sake of notation, we start numbering our terms with $i = 0$ as opposed to $i = 1$ as is usually done with series. This is helpful because then the index matches the exponent for the variable. A general polynomial is therefore written as $a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \dots a_kx^k$.

Most polynomials will not have a nice sequence of a_0, a_1, a_2, etc., as the coefficients in a polynomial rarely follow a pattern that can be represented in a mathematical formula. So, although the sequence a_0, a_1, a_2 … consists of "random" values, we can still think of a polynomial as a series. If it can be thought of as a series, we can write it in sigma form.

SIGMA REPRESENTATION OF A POLYNOMIAL

A polynomial of degree k can be written in sigma form as $\sum_{i=0}^{k} a_ix^i$.

Again, if this is confusing, writing out a few terms will help us see what is going on.

When $i = 0$, $a_ix^i = a_0 \cdot x^0$, so the first term is a_0.

When $i = 1$, $a_i x^i = a_1 \cdot x^1$, so the second term is a_1x.

When $i = 2$, $a_i x^i = a_2 \cdot x^2$, so the third term is a_2x^2.

When $i = k$, $a_i x^i = a_k \cdot x^k$, so the k^{th} term is $a_k x^k$.

Since this is a series, we add these terms together, and our result is $a_0 + a_1x + a_2x^2 + \dots + a_k x^k$, our generic polynomial.

2.2.1 ADDING AND SUBTRACTING POLYNOMIALS

When polynomials are added and subtracted, we use an idea that in high school mathematics is generally called "combining like terms," a particularly vague and potentially confusing phrase. What exactly is a "like term"? How do we combine them? Why do we only combine like terms, and what is wrong with combining unlike terms?

These ideas are usually not given adequate or rigorous answers in most high school math courses, and so we will attempt here to address these questions mathematically using the series representation of a polynomial.

Suppose we have two different arrays of numbers that we wish to add together, like two different rows in a spreadsheet. Adding the rows together should result in a new row where each entry in the row consists of the sum of the corresponding entries in the original rows. The first entry should be the sum of the two original first entries, the second entry should be the sum of the two original second entries, and so on. If we are adding two sequences together, the same idea should apply.

EXAMPLE 2.2A: What sequence results if the following two sequences are added: (3, 7, 11, 15, 19, …) + (1, 4, 9, 16, 25, …)?

SOLUTION: The new sequence will be 4, 11, 20, 31, 44, … .

EXAMPLE 2.2B: What polynomial results if the following two polynomials are added: $(3 + 7x + 11x^2 + 15x^3 + 19x^4 + \dots) + (1 + 4x + 9x^2 + 16x^3 + 25x^4 + \dots)$?

SOLUTION: The new polynomial will be $4 + 11x + 20x^2 + 31x^3 + 44x^4 + \dots$.

Thought of this way, the coefficients of the polynomial are of central importance, representing the terms in a sequence. The variables raised to different powers are of secondary importance, acting more like placeholders or separators. When we add two polynomials, the coefficients of the x^3 terms are added together because these coefficients are in the same place in the sequence (the fourth term, to be precise). This is why it is helpful to have both polynomials written in the same form, either lowest power of x to highest power of x or vice versa. If a polynomial does not have a term from a particular power, sometimes we write a term with a coefficient of 0 to avoid potential confusion.

We can also write the sum of these polynomials in sigma form.

EXAMPLE 2.2C: Write the following polynomial sum in sigma form: $(3+7x+11x^2+15x^3+19x^4+\ldots)+(1+4x+9x^2+16x^3+25x^4+\ldots)$.

SOLUTION: The first series can be written as $\sum_{i=0}^{\infty}(3+4i)\cdot x^i$, and the second can be written as $\sum_{i=0}^{\infty}(i+1)^2\cdot x^i$, so it should not surprise us that the sum can be written as $\sum_{i=0}^{\infty}[3+4i+(i+1)^2]\cdot x^i$.

Since the index starts and ends at the same value, the formulas for the terms of series can be added directly since the terms in the new series are precisely the sum of the terms of the original series. Of course, this can only be done with polynomials whose coefficients are generated by a nice formula, which is not often the case.

EXAMPLE 2.2D: Subtract the following polynomials: $(6+2x-3x^2+x^4)-(3x+5x^2-7x^3)$.

SOLUTION: Note that both of these polynomials have terms that are "missing," meaning the coefficient for some terms is 0. Including these terms gives us $(6+2x-3x^2+0x^3+x^4)-(0+3x+5x^2-7x^3+0x^4)$. Imagining these polynomials as sequences and then combining the terms in the same position in the sequence yields $6-x-8x^2+7x^3+x^4$.

Adding and subtracting polynomials is relatively straightforward; we will turn our attention next to multiplying polynomials.

2.2.2 MULTIPLYING POLYNOMIALS

We will assume the reader has some familiarity with multiplying or "expanding" polynomials from high school Algebra 1 and Algebra 2. Most high school students learn how to multiply polynomials by relying on a strange algorithm called "FOIL" (an acronym for **F**irst terms, **O**utside terms, **I**nside terms, **L**ast terms), which only works for certain small degree polynomials. This thought process is inefficient and does not translate well to general polynomials.

Some high school students are familiar with the distributive property and may multiply out larger-degree polynomials by repeatedly using the distributive property and just "keep 'FOIL'-ing." This ignores the fact that "FOIL" only works when we have two polynomials to multiply together and each of these polynomials has only two terms. These ways of multiplying polynomials are either limited, slow, or both.

In this section, we will try to improve upon these methods by continuing to focus on the coefficients of our polynomials rather than on the powers of the variable. We will begin with multiplying two polynomials that both have two terms.

Example 2.2e: What is $(x - 3) \cdot (2x + 1)$?

Solution: This is a classic "FOIL" problem, or perhaps a double-distribution problem. Most students in high school mathematics will write out:

$(x - 3) \cdot (2x + 1)$

$2x^2 + x - 6x - 3$

Then we "combine like terms" to get $2x^2 - 5x - 3$.

All well and good, but very slow, and it doesn't extend to harder problems very well.

Example 2.2f: What is $(x^2 + 11x - 2) \cdot (3x^3 - 16x^2 + 1)$?

Solution: Our original problem had two terms in each polynomial, so there were a total of four terms. Our new problem will have nine terms, and then these terms must be combined. The large number of calculations makes this method very slow, and also makes it more likely we will make a mistake. Let's return to our original problem and see if there is a nicer way to perform this calculation.

EXAMPLE 2.2E (Revisited): What is $(x - 3)\ (2x + 1)$?

SOLUTION: Rather than focusing on the terms in the given polynomials, let's think about the polynomial that is our answer. What will the highest power of x be in this polynomial? The highest power of x will be 2, and it will come from $x \cdot 2x$, so we will have $2x^2$. Will there be any terms of just x? Yes. These will come from $x \cdot 1$ and $-3 \cdot 2x$, so we will have a total of $-5x$. Will there be a constant term? Yes, this will come from $-3 \cdot 1$, so it will be -3. Therefore, our answer is $2x^2 - 5x - 3$.

This may not seem faster now, but this idea extends to larger degree polynomial multiplication problems and makes them much easier.

EXAMPLE 2.2G: What will the coefficient of x^4 be in $(x^2 + 11x - 2) \cdot (3x^3 - 16x^2 + 1)$?

SOLUTION: Rather than focusing on the result of each individual distribution or "FOIL," we will focus on the form of the answer and group coefficients rather than "like terms." In this product, there will be two terms of x^4: one from $x^2 \cdot -16x^2$, and one from $11x \cdot 3x^3$, giving us a total of $17x^4$, so the coefficient will be **17**.

EXAMPLE 2.2F (Revisited): What is $(x^2 + 11x - 2) \cdot (3x^3 - 16x^2 + 1)$?

SOLUTION: The highest power of x will be x^5, which will result from $x^2 \cdot 3x^3$, for $3x^5$. This will be the only term that generates x^5.

As we saw above, the coefficient for x^4 will be 17.

The terms that will generate x^3 are $11x \cdot -16x^2$ and $-2 \cdot 3x^3$, for a total of $-182x^3$.

The terms that will generate x^2 are $x^2 \cdot 1$ and $-2 \cdot -16x^2$, for a total of $33x^2$.

The only term that will generate x is $11x \cdot 1$, which is $11x$.

And, the only constant term will clearly be -2. Therefore, the product of these two polynomials is $3x^5 + 17x^4 - 182x^3 + 33x^2 + 11x - 2$.

A few remarks on this method or procedure. First, the Multiplication Principle tells us there should be $3 \cdot 3 = 9$ total terms to consider, and if desired we can keep track of the number of terms during the multiplication to make sure nothing was missed. In Example 2.2f, the number of terms for each power of x was 1, 2, 2, 2, 1, and 1, which sum to 9 as desired. Indeed, this use of the Multiplication Principle points out that each term in the product is generated by selecting one term from the first polynomial and one term from the second polynomial. This idea will become very important in the next part of the *Mathematics Resource Guide*. Second, once practiced, this method is faster and more accurate than "FOIL"-ing or distributing out all terms and then combining them. If mistakes are made, they are easier to find and fix. This method is mostly a manner of shifting the focus from the variable to the coefficients. We will conclude this section with one slightly more complicated example.

Example 2.2h: What is $(x - 3) \cdot (x^2 + 5x + 1) \cdot (2x - 5)$?

Solution: This product should have a total of $2 \cdot 3 \cdot 2 = 12$ terms, so we will keep track of these as we go to make sure we didn't miss anything.

The highest power of x will be x^4, which will come from $x \cdot x^2 \cdot 2x$, for $2x^4$.

The x^3 term will come from $x \cdot x^2 \cdot -5$, $x \cdot 5x \cdot 2x$, and $-3 \cdot x^2 \cdot 2x$, for a total of $-5 + 10 - 6 = -x^3$.

The x^2 term will come from $x \cdot 5x \cdot -5$, $x \cdot 1 \cdot 2x$, $-3 \cdot x^2 \cdot -5$, and $-3 \cdot 5x \cdot 2x$, for a total of $-25 + 2 + 15 - 30 = -38x^2$.

The x term will come from $x \cdot 1 \cdot -5$, $-3 \cdot 5x \cdot -5$, and $-3 \cdot 1 \cdot 2x$, for a total of $-5 + 75 - 6 = 64x$.

The constant term will be 15, and so the product of these three polynomials is $2x^4 - x^3 - 38x^2 + 64x + 15$.

The number of terms for the powers of x were 1, 3, 4, 3, and 1, for a total of 12, as expected.

In the next section, we will consider multiplying a polynomial by itself many times, or raising it to various powers. This will lead us to the Binomial Expansion Theorem.

2.3 THE BINOMIAL EXPANSION THEOREM

In the previous section, we studied multiplying polynomials. In this section, we will look at a specialization of multiplying polynomials: raising a polynomial to a power. In order to keep our discussion somewhat limited in scope, we will look at only **binomials**, or polynomials with two terms. The generalization of taking a binomial to a power is called the **Binomial Expansion Theorem**.

Let's start with a simple binomial: $(x + 1)$. First, $(x + 1)^2 = x^2 + 2x + 1$, so let's move to $(x + 1)^3$.

EXAMPLE 2.3A: What is the expanded form of $(x + 1)^3$?

SOLUTION: So we're sure we don't miss anything, we'll write it out as $(x + 1) \cdot (x + 1) \cdot (x + 1)$. The highest power of x is x^3, which comes from $x \cdot x \cdot x$, which is x^3 . The x^2 term will come from $x \cdot x \cdot 1$, $x \cdot 1 \cdot x$, and $1 \cdot x \cdot x$, for $3x^2$. The x term will come from $x \cdot 1 \cdot 1$, $1 \cdot x \cdot 1$, and $1 \cdot 1 \cdot x$, for $3x$. The constant term will be 1, so $(x + 1)^3$ expands to $x^3 + 3x^2 + 3x + 1$.

That seemed straightforward enough. Let's try a higher power.

EXAMPLE 2.3B: What is the expanded form of $(x + 1)^{10}$?

SOLUTION: So we can keep track of everything, we'll write:

$$(x + 1)^{10} = (x + 1) \cdot (x + 1) \cdot (x + 1) \cdot (x + 1) \cdot (x + 1) \cdot (x + 1) \cdot (x + 1) \cdot (x + 1) \cdot (x + 1) \cdot (x + 1)$$

The highest power of x is clearly x^{10}, and this only comes from $xxxxxxxxxx$, so there is only one x^{10}. The x^9 terms come from selecting the x from every $(x + 1)$ except one, so there should be 9 ways to do this, for $9x^9$.

But what about x^8? To generate an x^8, we need to pick the x from eight terms and the 1 from two terms, like *xxxxxxxx11*. But how many ways is this possible? Listing them all out seems extremely inefficient and time consuming: *xxxxxxx1x1, xxxxxx1xx1*, … . How many ways can this happen? We have ten spaces, and we need to select 8 of them to place an x. We can't select a place more than once, and the order of selection does not matter. This sounds like a job for…combinations!

So, there should be $\binom{10}{8} = 45$ ways to select eight x's and two 1's from the ten terms of $(x + 1)$, so the x^8 term will be $45x^8$.

Will this happen with every power of x? To generate x^7, we need seven x's and three 1's, like *xxxxx1x1x1*. How many ways can this happen? With ten spots, we need to select seven of them as x, we cannot select a spot more than once, and the order of selection does not matter. So, there are indeed $\binom{10}{7} = 120$ ways to create an x^7, which means the x^7 term will be $120x^7$.

Therefore, there are $\binom{10}{k}$ ways to generate a term of x^k, so each term in the expansion of $(x + 1)^{10}$ will be of the form $\binom{10}{k} \cdot x^k$. Writing this out while leaving the coefficients in terms of combinations yields:

$$(x+1)^{10} = \binom{10}{10} \cdot x^{10} + \binom{10}{9} \cdot x^9 + \binom{10}{8} \cdot x^8$$

$$+ \binom{10}{7} \cdot x^7 + \binom{10}{6} \cdot x^6 + \binom{10}{5} \cdot x^5 + \binom{10}{4} \cdot x^4 + \binom{10}{3} \cdot x^3 + \binom{10}{2} \cdot x^2 + \binom{10}{1} \cdot x^1 + \binom{10}{0} \cdot x^0.$$

This is, of course, a bit cumbersome to write, so writing this polynomial in sigma form makes sense. Therefore, $(x+1)^{10} = \sum_{k=0}^{10} \binom{10}{k} \cdot x^k$.

This is a very nice generalization. Let's try another expansion like this, but one that is a little more complicated, and see if the same type of mathematical structure applies.

Example 2.3c: What is $(2x + 5)^{15}$?

Solution: Rather than write out all fifteen terms of $2x + 5$, we'll imagine them written out in a long line. As in Example 2.3b, each term in the product is generated by selecting either $2x$ or 5 from each of the fifteen terms and multiplying them together.

So, $(2x) \cdot 5 \cdot (2x) \cdot (2x) \cdot (2x) \cdot 5 \cdot (2x) \cdot (2x) \cdot 5 \cdot (2x) \cdot (2x) \cdot (2x) \cdot 5 \cdot (2x) \cdot (2x)$ is one possible term in the expansion, and this term would be grouped with the other x^{11} terms. Let's go through a few powers of x and try to determine the coefficients.

The highest power of x will be x^{15} , and this will occur when the $2x$ is selected from all fifteen terms. $(2x)^{15} = 2^{15} \cdot x^{15}$, and so the coefficient of x^{15} will be 2^{15}, or 32,768.

The next highest power of x is x^{14} , and these will be generated by selecting the $2x$ from every term except one, from which a 5 will be selected. Each of these terms will therefore look like:

$(2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot 5$,

and so each will equal $(2x)^{14} \cdot 5 = 2^{14} \cdot 5 \cdot x^{14}$, or $81920 \cdot x^{14}$. But how many of these terms will be created? With fifteen spots and needing to select 14 of them as $2x$ (or one of them as 5), there should be 15 ways to do this. Note that $\binom{15}{14} = \binom{15}{1} = 15$, because selecting 14 spots for the $2x$ automatically selects the spot for the 5, and vice versa. Therefore, the final coefficient of x^{14} will be $81{,}920 \cdot 15 =$ 1,228,800.

Sometimes the numerical values disguise rather than assist generalization. Instead of thinking about this as $1228800 \cdot x^{14}$, let's keep track of where it came from: $\binom{15}{14} \cdot (2x)^{14} \cdot 5$. Again, this calculation makes sense: when selecting $2x$ from 14 of the 15 terms, $2x$ will be raised to the 14th power, and 5 will be selected from the last term. This selection can then occur in $\binom{15}{14}$ ways, since there are 15 terms and we are selecting the $2x$ from 14 of them.

What about x^{13} ? These are generated by selecting the $2x$ from 13 of the 15 terms. For example:

$(2x) \cdot 5 \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot 5 \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x) \cdot (2x)$. Each of these terms will therefore equal $(2x)^{13} \cdot 52$. But how many of these terms will there be? With 15 spots, selecting

13 of them to be $2x$ needs to occur without replacement, and the order of selection does not matter. So this can happen in $\binom{15}{13}$ ways, for a final answer of $\binom{15}{13}(2x)^{13}\cdot 5^2$.

It seems this pattern will continue for each power of x throughout our expansion. The x^4 term should come from $\binom{15}{4}(2x)^4\cdot 5^{11}$. (Convince yourself this is true if you are unsure!) Therefore, we can write out the product this way:

$$(2x+5)^{15} = \binom{15}{15}\cdot(2x)^{15}\cdot 5^0 + \binom{15}{14}\cdot(2x)^{14}\cdot 5^1 + \binom{15}{13}\cdot(2x)^{13}\cdot 5^2 + \ldots$$
$$+\binom{15}{2}\cdot(2x)^2\cdot 5^{13} + \binom{15}{1}\cdot(2x)\cdot 5^{14} + \binom{15}{0}\cdot(2x)^0\cdot 5^{15}.$$

This is, of course, long and tedious. Shouldn't there be a nicer way to write this?

How did we know how to write down each term in the form above? The $2x$ was raised to a power, 5 was raised to the power of 15 minus the power of $2x$, and both of these were multiplied by 15 choose the power of $2x$. Rather than saying "the power of $2x$," if we say k, we are then able to write this polynomial in sigma form:

$$(2x+5)^{15} = \sum_{k=0}^{15}\binom{15}{k}\cdot(2x)^k\cdot 5^{15-k}.$$

Let's make sure we understand this because, as we can hopefully see, this is an extremely powerful notational move.

When $k=0$, the term is $\binom{15}{0}\cdot(2x)^0\cdot 5^{15}$, which is the last term in the expansion as written out above.

When $k=1$, the term is $\binom{15}{1}\cdot(2x)^1\cdot 5^{14}$, which is the second to last term in the expansion as written out above.

When $k=4$, the term is $\binom{15}{4}\cdot(2x)^4\cdot 5^{11}$, which is what was predicted above.

When $k=14$, the term is $\binom{15}{14}\cdot(2x)^1\cdot 5^{14}$, which is the second to last term in the expansion as written out above.

So, other than listing the terms from lowest power of x to highest power of x, this form is identical to the expansion written out term by term (the long way). Based on our examples, it seems we are ready to generalize.

THE BINOMIAL EXPANSION THEOREM

When expanding a binomial in the form $(x + y)^n$, each term in this expansion is of the form $\binom{n}{k} \cdot x^k \cdot y^{n-k}$. Therefore, $(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^k \cdot y^{n-k}$.

Before we use this theorem creatively, let's make sure we can use it in the traditional sense: finding coefficients of the expansion of binomials raised to powers.

EXAMPLE 2.3D: What is the coefficient of z^6 in the expansion of $(z - 3)^9$?

SOLUTION: First we will walk through this problem as we did with our two previous examples—without using the binomial expansion theorem directly. Then we will use the theorem. (And hopefully reach the same answer!)

In the expansion of $(z - 3)^9$, each term is generated by selecting either z or -3 from each of the nine terms. To get a term with z^6, z must be selected from six terms and -3 selected from three terms. For example, $(-3) \cdot z \cdot z \cdot z \cdot (-3) \cdot z \cdot z \cdot (-3) \cdot z$ is one such term. Therefore, each term will be $-27z^6$. (Why will the coefficient be negative?) But how many of these terms will be created? With nine spots and needing to select six of them for z, this term can be built in $\binom{9}{6} = 84$ different ways. Therefore, the coefficient of z^6 will be $84 \cdot (-27) =$ **−2268**.

We used this thought process when we were developing the Binomial Expansion Theorem. Using the theorem directly, then, should result in the same calculation.

As $(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^k \cdot y^{n-k}$, substituting z for x, -3 for y, and 9 for n yields:

$(z - 3)^9 = \sum_{k=0}^{9} \binom{9}{k} \cdot z^k (-3)^{9-k}$. The z^6 term will occur when $k = 6$, so the term will be $\binom{9}{6} \cdot z^6 \cdot (-3)^3$, which is $\mathbf{-2268z^6}$, as above.

Of course, if the Binomial Expansion Theorem was only used for expanding binomials, it wouldn't be that interesting or useful. Later in the *Mathematics Resource Guide*, we will see how the Binomial Expansion Theorem is connected to the Binomial Distribution, an important topic in statistics. Because of its connections to combinations, the Binomial Expansion Theorem is also important in the study of combinatorics. We will conclude with one introductory example of this use of the Binomial Expansion Theorem.

EXAMPLE 2.3E: Prove $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n-1} + \binom{n}{n} = 2^n$ for all n.

SOLUTION: There are many ways to prove this classic result, and although no one method is "more correct" than another, the use of the Binomial Expansion Theorem is particularly nice.

To see how we might think to use the Binomial Expansion Theorem, let's start by writing the sum of the combinations in sigma form: $\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n-1} + \binom{n}{n} = \sum_{k=0}^{n} \binom{n}{k}$. This looks promising, as it is part of the Binomial Expansion Theorem, $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^k \cdot y^{n-k}$. We want the combinations portion of the sigma expression, but not the x^k or y^{n-k}. How can we get these terms to not be important anymore, or not impact the right-hand side of the equation? The combinations are multiplied by these terms, so if we want them to go away, maybe we can let $x = 1$ and $y = 1$. Then the right-hand side of the equation will be what we want, just $\sum_{k=0}^{n} \binom{n}{k}$, since 1 to any power is always 1.

But wait! Substituting $x = 1$ and $y = 1$ into the Binomial Expansion Theorem gives us what we need! $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^k \cdot y^{n-k}$, so after the substitution, we have $(1+1)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot 1^k \cdot 1^{n-k}$, or $2^n = \sum_{k=0}^{n} \binom{n}{k}$, and we are done.

As stated, we will return to the Binomial Expansion Theorem when we discuss the Binomial Distribution in the statistics portion of the *Mathematics Resource Guide* (strange that those would be related…). For now, we will shift our focus to a topic most people will interact with at some point in their lives: borrowing and investing money.

2.4 COMPOUND INTEREST

When we loan money to our friends or family, usually we don't charge them interest. Even if we do charge them interest, it is probably in the form of a fixed fee per time period, say $5 every week. This type of interest is called **simple interest** and is not how banks charge interest. When banks charge interest, interest is charged on the interest previously charged. This type of interest is called **compound interest**. Unfortunately, this type of interest results in more money being owed, especially for larger amounts of money borrowed over a longer period of time. Fortunately, bank accounts accrue interest in the same way, so that more money is earned for larger amounts of money or if money is invested over a longer period of time. We will begin by looking at a few examples and will then generalize a formula for compound interest.

2.4.1 INVESTING AND BORROWING

When we say that a bank account earns 2.5% interest, what we mean is that after a period of time (most banks calculate interest monthly or quarterly), 2.5% of the amount of money currently in the account is added to the account. When we are dealing with a compound interest situation, the interest previously earned is included the next time compounding occurs. This interest is then, in turn, included in the amount the next time interest is added. It seems that a recursive formula may be a good place to start.

EXAMPLE 2.4A: A bank account starts with \$500 and earns 2.5% interest each year. Write a recursive formula for the amount of money in the bank account after *t* years.

SOLUTION: Since the bank account has \$500 to start, $a^0 = \$500$. We'll use a_0 because there is \$500 after "0 years," so the index will match the number of years that have passed. Since the account earns 2.5% interest each year, we add 2.5% of the amount of money previously in the account to get to the next year's amount: $a_i = a_{i-1} + (0.025)a_{i-1}$. Noticing that both of these terms contain a_{i-1} allows us to simplify this to $a_i = 1.025 \cdot a_{i-1}$. Therefore, our recursive formula is $\mathbf{a_0 = \$500;\ a_i = 1.025 \cdot a_{i-1}}$.

Writing out the terms of this sequence yields $a_0 = \$500$, $a_1 = \$512.50$, $a_2 = \$525.31$, $a_3 = \$538.45$, and so on. Is there a direct formula for this sequence?

Yes! Since this sequence was formed by multiplying to get from one term to the next, this is a geometric sequence. We recall that a geometric sequence with the recursive formula $x_1 = a$; $x_i = r \cdot x_{i-1}$ has the direct formula $x_k = a \cdot r^{k-1}$. However, we began our index in this case with 0 as opposed to 1, so $x_k = a \cdot r^k$. (Think about why this is true—now we need to multiply by r once to get to x_1, twice to get to x_2, and so on.)

Therefore, the direct formula for this bank account is $500 \cdot (1.025)^t$, where t represents the number of years the money has been in the account.

Let's try another example and see if the same general approach applies.

EXAMPLE 2.4B: Frank borrows $750 from his local bank, and he is charged 3% annual interest, but interest is compounded every quarter. Assuming Frank does not repay any money along the way, how much money will Frank owe in three years?

SOLUTION: First we need to know one additional piece of information about how banks advertise and charge interest rates. Most banks advertise the annual percentage rate, or APR, of their accounts and loans. Interest is compounded more frequently, most often monthly or quarterly (four times a year), and so the interest is divided out equally over the whole year. So, in this example, Frank is charged .75% interest four times per year.

At each compounding, then, the amount Frank owes is multiplied by 1.0075. Over three years, interest will be compounded 12 times, so the final amount he owes is $750 \cdot (1.0075)^{12} =$ **$820.36**.

Is there really a difference between being charged 3% interest annually for three years and .75% quarterly for three years? If Frank was charged 3% interest annually for three years, he would owe $750 \cdot (1.03)^3 =$ $819.55. Although the difference in this case is very small, we can imagine that over longer periods of time and with larger amounts of money, this difference could become much greater.

On the basis of these examples, it seems we are ready to generalize.

COMPOUND INTEREST FORMULA

When P dollars is invested (or borrowed) in a situation using compound interest with interest rate r earned (or charged) at each compounding, after k compoundings, the value is $P \cdot (1 + r)^k$. To reflect the way banks divide the annual interest over the entire year, this formula is sometimes written as $P \cdot \left(1 + \frac{r}{n}\right)^{nt}$, where r represents the annual percentage interest rate, n represents the number of compoundings per year, and t represents the number of years.

This direct formula is built exactly as a direct formula for a geometric sequence is built: $x_k = a \cdot r^{k-1}$. Recall that r in a geometric sequence represents the ratio that is multiplied by to get from one term to the next.

Here r represents the annual interest rate for the account, so $1 + r$ becomes the ratio between consecutive terms. This is because the bank does not take away the money in your account when you earn interest; the bank leaves it there and adds the interest earned by your money: $a_i = a_{i-1} + r \cdot a_{i-1}$. The idea that this can be rewritten algebraically as $a_i = a_{i-1} \cdot (1 + r)$ is crucial, so this sequence can be identified as geometric.

Let's try two more examples to solidify our understanding of this process.

EXAMPLE 2.4C: Jill borrows $1,250 at 4% annual interest compounded monthly. Assuming she does not pay back any money in the meantime, how much money will Jill owe at the end of 5 years?

SOLUTION: As Jill's account accrues interest each month, it will compound 60 times in 5 years, and her 4% interest will be divided into $\frac{.04}{12}$ =0.333% interest each month. Therefore, using $P \cdot \left(1 + \frac{r}{n}\right)^{nt}$ results in $1250 \cdot \left(1 + \frac{.04}{12}\right)^{60} = 1526.25$, so Jill will owe a total of **$1,526.25** at the end of five years.

EXAMPLE 2.4D: Marco invests $2,000 in a certificate of deposit (CD) that earns 3.75% interest compounded quarterly. When Marco goes back to the bank and collects his money in 4 years, how much will his CD be worth?

SOLUTION: Marco's CD earns interest quarterly, so 4 times per year. This means his CD will have earned interest a total of 16 times, so $2000 \cdot \left(1 + \frac{.0375}{4}\right)^{16} =$ **$2,322.05**. Marco will have made about $322 on his investment.

2.4.2 ANNUITIES AND LOANS

Although sometimes people invest money as Marco did in the previous example, oftentimes people do not have a sum of money that they are willing to put in an account or CD for 5 or 10 years. More often, people are willing to invest a smaller amount of money into an account at regular intervals; say, $50 or $100 a month. In terms of borrowing, borrowing a large amount of money to buy a car or house, waiting 5 years or 30 years for all of the interest to accrue, and then suddenly having to pay back a massive amount of money is not feasible for most people. When we borrow money for a car or a house, we would like to pay back that money slowly over time, by paying a certain amount each month. In this part of the

Mathematics Resource Guide we will investigate these types of questions. By the end, you will understand the mathematical ideas and formulas behind annuities (investing money repeatedly over time) and loans (borrowing large amounts of money and then paying them back slowly over time).

EXAMPLE 2.4E: Megan invests in an annuity that earns 2.5% interest compounded monthly. If Megan deposits \$50 per month (and does not withdraw any money from her account), how much money will she have after 15 years?

SOLUTION: This problem is potentially more complicated than it looks. Megan will deposit a total of $\$50 \cdot 12 \cdot 15 = \$9{,}000$ into the account, but $9000 \cdot \left(1+\frac{.025}{12}\right)^{180}$ won't give the correct amount of money because not all of the \$9,000 is in the account for the entire 15 years. Only the first \$50 initially deposited is in the account for all 180 months, which means it is compounded 179 times. The \$50 that is deposited in the second month is compounded 178 times, and the \$50 that is deposited at the beginning of the third month is compounded 177 times, and so on until the last \$50 that is not compounded at all. Therefore, the total amount of money Megan has after the fifteen years can be written as:

$$50 \cdot \left(1+\frac{.025}{12}\right)^{179} + 50 \cdot \left(1+\frac{.025}{12}\right)^{178} + 50 \cdot \left(1+\frac{.025}{12}\right)^{177} + \ldots$$

$$\ldots + 50 \cdot \left(1+\frac{.025}{12}\right)^{2} + 50 \cdot \left(1+\frac{.025}{12}\right)^{1} + 50 \cdot \left(1+\frac{.025}{12}\right)^{0}.$$

This is, of course, somewhat awkward and long, but we notice a few nice things about this sum. One is that it can be written in sigma form as $\sum_{k=0}^{179} 50 \cdot \left(1+\frac{.025}{12}\right)^{k}$. The second is that to move from one term to the next (when written in order of increasing powers as opposed to decreasing), we multiply by $\left(1+\frac{.025}{12}\right)$. Therefore, this is a geometric series!

The formula for the sum of a geometric series is $S = \frac{a \cdot r^k - a}{r-1}$, where a is the first term, r is the common ratio, and k is the number of terms. In this case, $a = 50$, $r = \left(1+\frac{.025}{12}\right)$, and $k = 180$, so substituting these into the formula gives us:

$$S = \frac{50 \cdot \left(1+\frac{.025}{12}\right)^{180} - 50}{\left(1+\frac{.025}{12}\right) - 1} = \frac{50 \cdot \left[\left(1+\frac{.025}{12}\right)^{180} - 1\right]}{\left(\frac{.025}{12}\right)}$$, which evaluates to **\$10,906.17**.

This is not bad, considering Megan invested a total of \$9,000. It is not as good as it would have been had she invested the \$9,000 at the beginning, as $9000 \cdot \left(1 + \frac{.025}{12}\right)^{180} = \$13{,}089.82$, but we don't all have \$9,000 that we are willing to invest for 15 years. A smaller payment on regular intervals, such as \$50 a month, seems more reasonable.

Let's see if we can look at the final calculation and try to determine where each part of the calculation came from in order to generalize to a formula. Where did each piece of $\dfrac{50 \cdot \left(\left(1 + \frac{.025}{12}\right)^{180} - 1\right)}{\left(\frac{.025}{12}\right)}$ originate in our problem? The \$50 is clearly the amount of money invested at each compounding, and the exponent of 180 is equal to the number of months money was invested into the annuity. $\frac{.025}{12}$ is the amount of interest earned at each compounding, as the 2.5% per year is divided evenly over each month. Therefore, it seems we are ready to generalize.

ANNUITY FORMULA

If A dollars is invested n times per year in an annuity earning r annual interest compounded n times per year, the value of the annuity after t years is given by the formula $\frac{A \cdot [(1+i)^{n \cdot t} - 1]}{i}$, where i is the interest earned at each compounding, so $i = \frac{r}{n}$.

EXAMPLE 2.4F: Walter invests \$1,000 a year in a retirement fund that earns 5% annual interest for 35 years. How much money does Walter have in his fund when he retires?

SOLUTION: Walter is investing in an annuity, and the total value of this annuity is given by $\frac{A \cdot [(1+i)^{n \cdot t} - 1]}{i}$. In this case $A = 1000$, $n = 1$, $t = 35$, and $i = r = .05$. Therefore, after 35 years Walter will have $\frac{1000 \cdot [(1 + .05)^{35} - 1]}{.05} =$ **\$90,320.31.**

Pretty good, considering Walter invested a total of \$35,000—he almost tripled his money.

EXAMPLE 2.4G: Barb invested some amount of money each month in an annuity earning 3.2% annual interest compounded monthly. At the end of 20 years, Barb's annuity was worth a little more than $33,500. How much money did Barb invest each month?

SOLUTION: The unknown in this situation is the amount of money invested each month, A in our formula. Since all other values are known, we are able to solve for this variable.

$$33500 = \frac{A \cdot \left[\left(1+\frac{.032}{12}\right)^{20\cdot 12} - 1\right]}{\left(\frac{.032}{12}\right)}.$$

Evaluating the right-hand side yields: $33500 = A \cdot (335.57479)$, so $A = 99.8287$, and Barb invested **$100** per month.

Again, this is a fairly good return on Barb's investment. She deposited a total of $100 \cdot 12 \cdot 20 = \$24{,}000$ and earned almost a 40% return on her money over the life of the annuity. Certainly Barb's investment would have been worth more money if she had deposited all $24,000 at the beginning and let it grow for the 20 years, since $24{,}000 \cdot \left(1+\frac{.032}{12}\right)^{240} = \$45{,}476.79$. But this difference is not as great as one might think, and most people don't have $24,000 they can invest for 20 years.

The idea of paying small amounts of money over time is how loans work as well. Most people borrow money from a bank for a car, a home, or a college education. How does a bank determine the monthly payment for a particular loan? This is a fairly important piece of information it seems most knowledgeable and informed borrowers of money should know.

Let's say you borrow some money from a bank to buy a new car, say $20,000. Since cars depreciate in value over time, most car loans are for 5 years. (If you aren't able to make payments and the bank has to repossess the car, they want this to happen when the car still has value). If the bank did not loan you the money, it could have invested the money in a compound interest account at a particular interest rate for five years. The $20,000 the bank is loaning you is therefore worth more than $20,000 at the end of five years. Your monthly payment essentially goes into a five-year annuity whose goal is to equal the amount of money the bank could have made by investing the $20,000 in a compound interest account. The monthly payment is calculated so that at the end of the five years, the value of the annuity is equal to the value of the compound interest account. Note that this compound interest account is really a hypothetical account because the bank is not actually investing the $20,000 in this account; the bank is loaning it to you to buy a car. Let's see how this works in an example.

EXAMPLE 2.4H: You borrow $20,000 to purchase a new car. Your interest rate is 6.5% annual interest compounded monthly, and you take out a five-year loan. What is your monthly payment?

SOLUTION: To the bank, the $20,000 you borrowed will be worth $20000 \cdot \left(1+\frac{.065}{12}\right)^{60} = \$27{,}656.35$ at the end of the five years. Therefore, you need to deposit money each month into an annuity that will be worth this amount of money at the end of the five years. So:

$27656.35 = \dfrac{A \cdot \left[\left(1+\frac{.065}{12}\right)^{60} - 1\right]}{\left(\frac{.065}{12}\right)}$, and $27656.35 = A \cdot (70.673967)$, so $A =$ **$391.32**.

This means you paid a total of ($391.32) · 12 · 5 = $23,479.20, so the bank made about $3,500 by loaning you the money. This is less than the mythical $7,600 they would make if they did not loan you the money and instead invested it in the (hypothetical) compound interest account. So, why would the bank loan you the money? One reason is that if they invest the money, they are responsible for making sure they get the return they are interested in—it takes work to find and maintain an investment that earns 6.5% annual interest compounded monthly. By loaning you the money, they do not have to do this work to maintain the account (assuming you make your loan payments).

Another, probably more important, reason is the fact that interest rates for loaning money are generally higher than interest rates for investing money. So, although the bank loans money out at 6.5% annual interest, they probably are not able to invest money (consistently) at this rate. If the bank is able to invest at a more modest 4% interest rate, the initial $20,000 is worth $20000 \cdot \left(1+\frac{.04}{12}\right)^{60} = \$24{,}419.93$, which is extremely close in value to the $23,479.20 you end up paying the bank for the loan. Depending on the difference between the interest rates for loaning and investing, it may be more profitable for the bank to loan money than to invest it. Furthermore, by collecting on your loan payment, the bank is collecting money each month (as you pay your loan) rather than waiting five years to collect the money as they would under an investment.

Let's look at another example, this time for a home loan, before we generalize.

EXAMPLE 2.41: Miguel borrows \$175,000 at 5% annual interest compounded monthly to buy a house. If Miguel takes out a 30-year mortgage, what will his monthly house payment be?

SOLUTION: To the bank, the \$175,000 Miguel borrows will be (hypothetically) worth $175000 \cdot \left(1+\frac{.05}{12}\right)^{360} =$ \$781,855.25 (yikes!) after 30 years. In order to pay this off, Miguel (hypothetically) invests in an annuity with 5% interest compounded monthly, and he needs this annuity to equal \$781,855.25 after 30 years. So:

$781855.25 = \dfrac{A \cdot \left[\left(1+\frac{.05}{12}\right)^{360} - 1\right]}{\left(\frac{.05}{12}\right)}$. Evaluating the right-hand side yields $781855.25 = A \cdot (832.258635)$, so $A = 939.4378$. Miguel's monthly mortgage payment will be **\$939.44**.

Over the life of the mortgage, Miguel will end up paying the bank $30 \cdot 12 \cdot (\$939.44) = \$338{,}198.40$, well short of \$781,855.25, which is referred to as the "future value" of his initial \$175,000. But \$338,198.40 is still significantly more than the amount of the initial loan. When a home is purchased, by law the buyer must be informed of the total cost of the mortgage over the life of the loan prior to closing on the house.

Based on these two examples, it seems we are ready to generalize this process.

PROCESS FOR DETERMINING A LOAN PAYMENT

When determining a loan payment, the future value of the money borrowed must equal the value of an annuity where the (hypothetical) investment and (hypothetical) annuity have the same interest rate and compound with the same frequency. If P dollars is borrowed at annual interest rate r for t years with payments made n times per year, the following equation must hold: $P \cdot (1+i)^{n \cdot t} = \frac{A \cdot [(1+i)^{n \cdot t} - 1]}{i}$ where A represents the monthly payment and $i = \frac{r}{n}$.

Although this formula can be rearranged and solved for A, the mathematical payoff for this is minimal. It is more efficient to remember the process of how to calculate a loan payment and use the previously established compound interest and annuity formulas.

EXAMPLE 2.4J: Sarah has been renting an apartment for the last year and is ready to buy a house. She feels she can afford a monthly mortgage payment of \$750. If current mortgage interest rates are 6% and she is interested in a 30-year loan, about how much money can Sarah expect to be able to borrow?

SOLUTION: Substituting into the our previously determined equation yields: $P \cdot \left(1+\frac{.06}{12}\right)^{360} = \dfrac{750 \cdot \left[\left(1+\frac{.06}{12}\right)^{360} - 1\right]}{\left(\frac{.06}{12}\right)}$. Calculating reduces this equation to $P \cdot (6.022575) = 753386.2818$, so $P = 125093.715$. Therefore Sarah can anticipate being able to borrow about **\$125,000** to purchase a house.

2.5 EULER'S CONSTANT

We will conclude this section of the *Mathematics Resource Guide* with a look at one of the most important numbers in mathematics: **Euler's constant, *e***. It is often referred to as the "natural number," due to one way it is determined, but *e* is related to and used in many mathematical contexts. We will not look at the importance of *e* in calculus here, but instead will see if we can connect it to combinations. We will begin, however, with a thought experiment that brings about *e*.

EXAMPLE 2.5A: Leonard invests \$1,000 in what seems like an investment opportunity that is too good to be true: 100% annual interest compounded monthly for 1 year. How much money will Leonard have when his unbelievable investment matures?

SOLUTION: Our first inclination is to think this will be, approximately, one hundred bazillion dollars. Let's see: compound interest is modeled by $P \cdot \left(1+\frac{r}{n}\right)^{nt}$, so after the 1 year, Leonard will have $1000 \cdot \left(1+\frac{1}{12}\right)^{12}$, which equals…**\$2,613.04!** Wait, what? That's not one hundred bazillion dollars? What happened?

EXAMPLE 2.5B: Leonard is still a bit puzzled about his previous investment of $1,000 at 100% annual interest compounded monthly. It seems like it should have been worth more money after 1 year. This time he invests $1,000 at 100% annual interest compounded daily for one year. How much is this investment worth?

SOLUTION: Okay, this time it's going to be one hundred bazillion dollars, right? 100% annual interest compounded *daily*! That's going to be ridiculous. $1000 \cdot \left(1+\frac{1}{365}\right)^{365}$ = **$2,714.57**. What! Only $100 more? That's crazy! We must not be compounding enough times.

EXAMPLE 2.5C: Leonard is getting more and more perplexed. This time he invests $1,000 in an account earning 100% interest compounded every second for one year. This time it is worth one hundred bazillion dollars, right?

SOLUTION: Unfortunately, $1000 \cdot \left(1+\frac{1}{31536000}\right)^{31536000}$ = $2,718.28, so Leonard makes "only" $1,718 on his investment of $1,000. From an investment standpoint, this is a very good return. But, from a mathematical standpoint, it seems strange that the total value of the investment is not increasing very much as the compounding rate increases. Indeed, although there are differences in the values of $1000 \cdot \left(1+\frac{1}{31536000}\right)^{31536000}$ and $1000 \cdot \left(1+\frac{1}{1000000000}\right)^{1000000000}$, in this investment context both of them result in the same amount of money: **$2,718.28**.

So what exactly is going on? Let's focus on the multiplier for the initial amount of money, $\left(1+\frac{r}{n}\right)^{nt}$. The interest rate is 100%, and the money is being invested for one year, so $r = 1$ and $t = 1$. We are therefore primarily concerned with $\left(1+\frac{1}{n}\right)^{n}$ for increasing values of n. Although the exponent is getting larger, the number being raised to this exponent is getting smaller. This creates an interesting mathematical power struggle between the base and the exponent. The table below shows the approximate value of $\left(1+\frac{1}{n}\right)^{n}$ for various values of n.

n	$\left(1+\frac{1}{n}\right)^{n}$
1	2
2	2.25
3	2.37037
5	2.48832
10	2.59374
100	2.7048138
1000	2.71692393
10000	2.718145926
1 million	2.7182804693
1 billion	2.7182818271
2 billion	2.7182818277

As n increases, it appears that $\left(1+\frac{1}{n}\right)^{n}$ does not approach infinity, but has a fixed number that it approaches. Mathematicians call this type of number an upper bound, and this is one way to define e: e is the number that $\left(1+\frac{1}{n}\right)^{n}$ approaches when n approaches infinity.

DEFINITION

Euler's constant, e, is the number that $\left(1+\frac{1}{n}\right)^{n}$ approaches as n increases to infinity. A decimal approximation of e is $e \approx 2.718281828$.

As this number arises from a compound interest context where n is the number of compoundings per year, e becomes important and useful when the number of compoundings per year equals infinity, or, in other words, when something is *compounded continuously*. Although very few (if any) bank accounts compound interest continually, many naturally occurring phenomena can be thought of as continuously compounding. Bacteria growing in a Petri dish, for example, don't compound on a set schedule ("All right everyone! Our two hours is up! Divide!), and modeling populations of different species of animals also works well if the population is assumed to compound continuously. Radioactive decay is another example of continuous compounding. Because e is used to model many of these naturally occurring activities, it is sometimes called the natural number.

But, so far all we can do is use e to model situations with 100% growth (interest) rate and continuous compounding. What if we want or need to use a growth rate other than 100%? How does this relate to e?

From our work with compound interest, we know that for any interest rate the multiplier will be $\left(1+\frac{r}{n}\right)^n$.

But, all we know is that $\left(1+\frac{1}{n}\right)^n$ approaches e as n approaches infinity. How can we turn $\left(1+\frac{r}{n}\right)^n$ into something like $\left(1+\frac{1}{n}\right)^n$?

First we need to recognize that in the second expression, n can act as a placeholder, and anything of the form $\left(1+\frac{1}{x}\right)^x$ will equal e as x approaches infinity. This means we can substitute anything we like in for x, and it will still equal e as x approaches infinity. Needing $\frac{r}{n}$ inside the parentheses, we notice that $\frac{1}{\left(\frac{n}{r}\right)}=\frac{r}{n}$, and so we substitute $\frac{n}{r}$ in for x. This gives us $e=\lim\limits_{n\to\infty}\left(1+\frac{1}{\left(\frac{n}{r}\right)}\right)^{\frac{n}{r}}$, or $e=\lim\limits_{n\to\infty}\left(1+\frac{r}{n}\right)^{\frac{n}{r}}$. Now the inside of the parentheses is what we need, which is good! We were trying to build the expression for compound interest, $\left(1+\frac{r}{n}\right)^n$. Raising both sides of this equation to the r power will give us what we need, and so $e^r=\lim\limits_{n\to\infty}\left(1+\frac{r}{n}\right)^n$. This equation says that the multiplier for continuous compounding at interest rate r is e^r. Often when we solve problems using continuous compounding, we will not know what the specific growth rate is, ("Excuse me, Mr. Bacteria, can you tell me the growth rate at which you are compounding? 6.4%? Great, thanks!) and will treat e^r as one number representing the multiplier. If the reader is familiar with logarithms, this equation contains extra mathematical significance.

This development is also significant in that it is the first time in this year's *Mathematics Resource Guide* that we have been able to use the structure of mathematics itself to answer a question about how to represent something mathematically. Previously, we looked at a situation, reasoned about some calculations, and then generalized to produce a formula or method. Here, however, it is extremely difficult to generalize from calculations that $e^r = \lim_{n \to \infty}\left(1+\frac{r}{n}\right)^n$, and instead we used mathematics directly to construct this formula. This formula, then, communicates some new knowledge about the physical world that we could not have known without mathematics.

Let's see how we might use this in a problem.

EXAMPLE 2.5D: 100 bacteria are placed in a Petri dish, and after three hours there are 185 bacteria. Assuming a continuous growth model, how many bacteria will there be in seven hours?

SOLUTION: First we need to construct a model for this situation. The bacteria are growing with some growth rate r, so the number of bacteria can be approximated by $\lim_{n \to \infty}\left[P \cdot \left(1+\frac{r}{n}\right)^{nt}\right]$. Since $e^r = \lim_{n \to \infty}\left(1+\frac{r}{n}\right)^n$, this expression becomes $P \cdot e^{rt}$, which hopefully looks familiar to those who have taken enough science classes. This expression equals 185 when $t = 3$, so $185 = 100 \cdot e^{3r}$ or $1.85 = e^{3r}$. We then take the cube root of each side to obtain $1.227601 = e^r$. If we have an understanding of logarithms, it is possible to solve for r, but that is not necessary to answer this question. We want to know how many bacteria there will be after seven hours, which is given by the expression $100 \cdot e^{7r}$. This can be rewritten as $100 \cdot (e^r)^7$, and as $e^r = 1.227601$, this expression evaluates as $100 \cdot (1.227601)^7 = 420.14645$. Therefore, after seven hours there will be about **420** bacteria.

We will conclude this section with a look at how e potentially relates to combinations, a key topic we explored in Section 1 of this resource guide. There are two major mathematical ways to represent e, and we started with one of them: $e = \lim_{n \to \infty}\left(1+\frac{1}{n}\right)^n$. We will use this representation to derive the other major representation of e. To get there, we will use the Binomial Expansion Theorem.

We will start by considering $\left(1+\frac{1}{n}\right)^n$. Once isolated, we notice that this is a binomial raised to a power. The Binomial Expansion Theorem therefore applies, and we will substitute $y = 1$ and $x = \frac{1}{n}$. (Since these terms are added, we can consider them in either order, and this substitution makes the powers work out a bit more nicely.) Formally, the Binomial Expansion Theorem says:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^k \cdot y^{n-k}$$

Substituting $y = 1$ and $x = \frac{1}{n}$ gives us: $\left(\frac{1}{n} + 1\right)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot \left(\frac{1}{n}\right)^k \cdot 1^{n-k}$.

Simplifying the right-hand side and rewriting the left-hand side yields:

$$\left(1+\frac{1}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot \frac{1}{n^k}$$

This sigma expression is somewhat difficult to follow, so let's write out a few terms.

When $k = 0$, $\binom{n}{k} \cdot \frac{1}{n^k}$ equals $\binom{n}{0} \cdot \frac{1}{n^0}$, so the first term is 1.

When $k = 1$, $\binom{n}{k} \cdot \frac{1}{n^k}$ equals $\binom{n}{1} \cdot \frac{1}{n^1} = n \cdot \frac{1}{n}$, so the second term is also 1.

When $k = 2$, $\binom{n}{k} \cdot \frac{1}{n^k}$ equals $\binom{n}{2} \cdot \frac{1}{n^2} = \frac{n \cdot (n-1)}{2} \cdot \frac{1}{n^2}$, so the third term is $\frac{1}{2} \cdot \frac{n \cdot (n-1)}{n^2}$.

When $k = 3$, $\binom{n}{k} \cdot \frac{1}{n^k}$ equals $\binom{n}{3} \cdot \frac{1}{n^3} = \frac{n \cdot (n-1) \cdot (n-2)}{3!} \cdot \frac{1}{n^3}$, so the fourth term is $\frac{1}{3!} \cdot \frac{n \cdot (n-1) \cdot (n-2)}{n^3}$.

When $k = k$, $\binom{n}{k} \cdot \frac{1}{n^k}$ equals $\binom{n}{k} \cdot \frac{1}{n^k} = \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{k!} \cdot \frac{1}{n^k}$, so the $(k+1)^{\text{th}}$ term is $\frac{1}{k!} \cdot \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{n^k}$.

These terms are all pretty ugly, but there is one important thing we have not yet done to $\left(1+\frac{1}{n}\right)^n$, which is where all this started. $e = \lim_{n \to \infty} \left(1+\frac{1}{n}\right)^n$, so let's consider what happens to each of these terms as n approaches infinity.

The first and second terms are 1, so as n approaches infinity, they still equal 1.

The third term is $\frac{1}{2} \cdot \frac{n \cdot (n-1)}{n^2}$, so as n approaches infinity, this term equals $\frac{1}{2}$.

The fourth term equals $\frac{1}{3!} \cdot \frac{n \cdot (n-1) \cdot (n-2)}{n^3}$, so as n approaches infinity, this term equals $\frac{1}{3!}$.

The $(k+1)^{\text{th}}$ term equals $\frac{1}{k!} \; \frac{n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot (n-k+1)}{n^k}$, so as n approaches infinity, this term equals $\frac{1}{k!}$.

Therefore, we can simplify all of this fairly drastically when we consider n approaching infinity.

$$e = \lim_{n \to \infty} \left(1+\frac{1}{n}\right)^n$$

$$= \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = \lim_{n \to \infty} \sum_{k=0}^{n} \binom{n}{k} \cdot \frac{1}{n^k}$$

$$= 1 + 1 + \frac{1}{2} + \frac{1}{3!} + \ldots + \frac{1}{k!} = + \ldots$$

So, we have reached a new way to write e: $e = 1 + 1 + \frac{1}{2} + \frac{1}{3!} + \ldots + \frac{1}{k!} = + \ldots$

Since $0! = 1$, $1! = 1$, and $2! = 2$, we can write all of the denominators in this series as factorials:

$$e = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{k!} + \ldots$$

This makes it possible to write the series definition of e: $e = \sum_{k=0}^{\infty} \frac{1}{k!}$.

Here we have a powerful example of mathematics generating more mathematics. In Euler's time, (he lived during the 1700s), approximating e using $\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$ would have been almost impossible, as $n = 1000$ only generates two decimal points of accuracy. Since $k!$ decreases so rapidly, using $\sum_{k=0}^{\infty} \frac{1}{k!}$ gives us much greater accuracy with far fewer terms and calculations that are possible to do by hand! Using $k = 6$, for example, gives us three decimal places of accuracy, and $k = 10$ gives us an approximation of e that is accurate to 7 digits!

Euler's constant is one of the most important numbers and mathematical concepts in the historical and intellectual development of mathematics. It helped bridge the gap between algebra and calculus and serves as an important cornerstone in mathematics.

We hope the reader has gained a greater appreciation of algebra as a branch of serious mathematical study—it is not just a collection of strange problems about solving equations or factoring. Algebra is much broader and more powerful than this and serves as the foundation for higher mathematics.

SECTION 2 SUMMARY: ALGEBRA

- **Definition of a Sequence:** A sequence is a list of objects presented in a particular order. The objects in the sequence are called the terms of the sequence.
- **Notation for the Terms of Sequences:** The position of a term in a sequence is called the *index* of the term. The terms of a sequence are denoted by a variable (usually a or x) and an index, with the index written as a subscript. Unless otherwise noted, the index begins at 1 and consists of positive counting numbers. A generic sequence will commonly be written as $x_1, x_2, x_3 \ldots$ or $a_1, a_2, a_3 \ldots$.
- **Notation for Sequences:** When the terms of a sequence can be generated from a direct formula, we can represent the sequence by giving the direct formula in curly brackets, such as $\{x_i = 4i - 1\}$. For a

sequence with a particular number of terms, we denote the starting index as a subscript and the ending index as a superscript, such as $\{x_i = 4i - 1\}_{i=1}^{25}$.

- **Recursive Formula:** A recursive formula for a sequence is a formula that declares the starting value (or values) for the sequence and how the subsequent terms are made from the previous term (or terms). For example, $a_1 = 5$; $a_i = 2 \cdot a_{i-1} + 3$.

- **Direct Formula:** A direct formula for a sequence is a formula that declares how to generate the terms of the sequence from the values of the index. For example, $\{x_i = i^2\}$.

- **Arithmetic Sequences:** An arithmetic sequence is a sequence with a constant difference between consecutive terms. The first term of an arithmetic sequence is usually represented by a, and the constant difference is represented by d.

- **Formulas for Arithmetic Sequences:** The recursive formula for an arithmetic sequence is $x_1 = a$; $x_i = x_{i-1} + d$. The direct formula for an arithmetic sequence is $x_k = a + (k - 1) \cdot d$.

- **Geometric Sequences:** A geometric sequence is a sequence with a constant ratio between consecutive terms. The first term for a geometric sequence is usually represented by a, and the constant ratio is represented by r.

- **Formulas for Geometric Sequences:** The recursive formula for a geometric sequence is $x_1 = a$; $x_i = r \cdot x_{i-1}$. The direct formula for a geometric sequence is $x_k = a \cdot r^{k-1}$.

- **Definition of a Series:** A series is the sum of the terms in a sequence.

- **Formula and Strategy for Arithmetic Series:** To find an arithmetic series, we create paired sums equal to the first term plus the last term. If there are k terms, there will be $\frac{k}{2}$ pairs, so the arithmetic series will be $(x_1 + x_k) \cdot \frac{k}{2}$. As $x_1 = a$ and $x_k = a + (k - 1) \cdot d$, this formula can be written using the first term and the constant difference as $[2a + (k - 1) \cdot d] \cdot \frac{k}{2}$.

- **Formula and Strategy for Geometric Series:** To find a geometric series, we call the series we are looking for S, add an additional term to S, and then algebraically manipulate. A geometric series with k terms and $r \neq 1$ is equal to $S = \frac{a \cdot r^k - a}{r - 1}$. Alternatively, we can write this formula by referencing the terms in the sequence: $S = \frac{x_{k+1} - x_1}{r - 1}$, where x_k is the last term in the series and x_{k+1} is the next term in the sequence (but is not included in the series).

- **Formula for Infinite Geometric Series:** An infinite geometric series equals a finite number if $x_k \approx 0$ for sufficiently large values of k (or, equivalently, $|r| < 1$). In this case, $S = \frac{a}{1 - r}$.

Sigma Notation: Mathematicians use sigma, $\sum$, to denote a series. $\sum_{i=a}^{b} f(i)$ means the series that corresponds to the sequence generated by the formula $f(i)$ where the index begins at a and ends at b. For example, $\sum_{i=1}^{5} i^2 = 1 + 4 + 9 + 16 + 25$.

Sigma Forms of Arithmetic and Geometric Series Formulas:

- *Arithmetic Series:* $\sum_{i=1}^{k} (a + (i-1) \cdot d) = [2a + (k-1) \cdot d] \cdot \frac{k}{2}$.
- *Finite Geometric Series:* $\sum_{i=1}^{k} (a \cdot r^{i-1}) = \frac{a \cdot r^k - a}{r-1}$.
- *Infinite Geometric Series:* $\sum_{i=1}^{\infty} (a \cdot r^{i-1}) = \frac{a}{1-r}$, when $|r| < 1$.

Definition of a Polynomial: A polynomial is an algebraic object consisting of terms. Each term is made up of a variable, usually x, raised to a different non-negative integer power and a coefficient. The highest power of x with a non-zero coefficient is called the *degree* of the polynomial.

Sigma Representation of a Polynomial: A polynomial of degree k can be written in sigma form as $\sum_{i=0}^{k} a_i x^i$.

Adding and Subtracting Polynomials: When polynomials are added or subtracted, the coefficients of the terms with the same power of the variable are added or subtracted.

Multiplying Polynomials: When polynomials are multiplied, one term from each polynomial in the product is selected. Each new term has its power equal to the sum of the powers of the old terms, and its coefficient equal to the product of the old coefficients.

The Binomial Expansion Theorem: When expanding a binomial in the form $(x + y)^n$, each term in this expansion is of the form $\binom{n}{k} \cdot x^k \cdot y^{n-k}$. Therefore, $(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} \cdot x^k \cdot y^{n-k}$.

Compound Interest: Compound interest occurs when the interest earned at each compounding is included in the calculation of interest at the subsequent compoundings. In other words, interest is earned on interest. Compound interest is a geometric sequence.

Compound Interest Formula: When P dollars is invested (or borrowed) in a situation using compound interest with interest rate r earned (or charged) at each compounding, after k compoundings the value is $P \cdot (1+ r)^k$. To reflect the way banks divide the annual interest over the entire year, this formula is sometimes written as $P \cdot \left(1 + \frac{r}{n}\right)^{nt}$, where r represents the annual percentage interest rate, n represents the number of compoundings per year, and t represents the number of years.

Annuities: An annuity is an account where money is repeatedly invested at regular intervals—for example, \$50 per month. Annuities are geometric series.

Annuity Formula: If A dollars is invested n times per year in an annuity earning r annual interest compounded n times per year, the value of the annuity after t years is given by the formula $\frac{A \cdot [(1+i)^{n \cdot t} - 1]}{i}$, where i is the interest earned at each compounding, so $i = \frac{r}{n}$.

Process for Determining a Loan Payment: When determining a loan payment, the future value of the money borrowed must equal the value of an annuity, where the (hypothetical) investment and the (hypothetical) annuity have the same interest rate and compound with the same frequency. If P dollars is borrowed at an annual interest rate r for t years with payments made n times per year, the following equation must hold: $P \cdot (1+i)^{n \cdot t} = \frac{A \cdot [(1+i)^{n \cdot t} - 1]}{i}$, where A represents the monthly payment and $i = \frac{r}{n}$.

Euler's Constant: Euler's constant is often used to model situations with continuous compounding. e has two major mathematical definitions: $e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n$ and $e = \sum_{k=0}^{n} \frac{1}{k!}$. A decimal approximation of e is $e \approx 2.718281828$.

Modeling Continuous Compounding: To model a situation with continuous growth (or decay), we use the expression $P \cdot e^{rt}$, where P represents the initial amount, r represents the rate of growth (or decay), and t represents the number of years that have passed.

SECTION 2 REVIEW PROBLEMS: ALGEBRA

1. Consider the following series: $41 + 43 + 45 + \ldots + 1203 + 1205 + 1207$.

   a. Write the sequence that corresponds to this series as a direct formula.

   b. Write the sequence that corresponds to this series as a recursive formula.

   c. Is this series arithmetic, geometric, or neither? Explain.

   d. Determine the number of terms in this series.

   e. Write this series in sigma notation.

   f. Find the value of this series.

2. Consider the following series: $2 - \frac{4}{5} + \frac{8}{25} - \ldots + \frac{2097152}{935367431640625} - \frac{4194304}{476837158203125}$

a. Write the sequence that corresponds to this series as a direct formula.

b. Write the sequence that corresponds to this series as a recursive formula.

c. Is this series arithmetic, geometric, or neither? Explain.

d. Determine the number of terms in this series.

e. Write this series in sigma notation.

f. Find the value of this series.

g. Find the value of this series if it had an infinite number of terms.

3. Consider the series 1 + 4 + 9 + 16 + 25 + 36 + 49 + … + 225 + 256.

   a. Write the sequence that corresponds to this series as a direct formula.

   b. Write the sequence that corresponds to this series as a recursive formula.

   c. Is this series arithmetic, geometric, or neither? Explain.

   d. Determine the number of terms in this series.

   e. Write this series in sigma notation.

   f. Create a new sequence such that the first term is the first term of the series above, the second term is the sum of the first two terms of the series, and in general the k^{th} term is the sum of the first k terms in the series.

   g. Describe this new sequence in words and write the k^{th} term of this sequence using sigma notation.

   h. Write a direct formula for this sequence.

4. Two employees in a grocery store are trying to figure out how to make a display using 140 cans of soup. They need to make a trapezoidal display so that each row has two fewer cans than the row below it. They have room for a total of 10 rows of cans before the display becomes unstable. How many cans should they place in the bottom row of the display?

5. A bouncy ball is dropped from a height of 24 feet. Each time the ball bounces, it returns to a height that is $\frac{2}{3}$ the previous height.

a. What height does the ball return to after the first bounce?

b. What height does the ball return to after the fourth bounce?

c. What is the total distance the ball travels?

6. The polynomial $11x^7 - 6x^6 + 2x^4 - x^3 + 16x + 15$ could be used to represent two different sequences. Write each of them.

7. Multiply the following polynomials: $(x^2 + 4x + 5) \cdot (2x - 3) \cdot (4x^3 + 2x - 1)$. Do not use a calculator.

8. What is the coefficient of x^5 in the expansion of $(3x + 2)^7$?

9. What is the coefficient of y^6 in the expansion of $(y^2 - 5)^7$?

10. Prove that $\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \ldots \pm \binom{n}{n-1} \ldots \binom{n}{n} = 0$. (Note that whether n is even or odd will determine the signs (+/–) of the last two terms.)

11. Sophie invested \$4,000 in an account earning compound interest compounded monthly. After 4 years, her account was worth \$5,750. What was Sophie's interest rate?

12. Peter deposits \$1,500 per year in a retirement account that earns 4.5% annual interest. After 40 years, how much money will be in Peter's account?

13. When Rachel is born, her parents begin investing money each month into an annuity for her college education that earns 4% annual interest. If Rachel's parents want to have \$120,000 when Rachel turns 20 years old, how much should her parents invest each month?

14. Juan borrows \$200,000 to purchase a house at 6.6% annual interest.

a. What is Juan's monthly mortgage payment, given that Juan takes out a 30-year mortgage?

b. Over the course of his mortgage, how much will Juan pay in mortgage payments to repay the loan?

15. Scientists have discovered a rare isotope of carbon, Carbon-14. An initial sample of 5.4 grams of Carbon-14 decays to 3.1 grams after 50 years. How much of this sample of Carbon-14 do scientists predict will remain after 130 years?

Section 3

Statistics

Compared to the larger field of mathematics, statistics is a very young field, having only been around for the last two hundred years or so. With an increase in scientific experimentation, people were in need of mathematics that would help them determine whether their results were caused by random chance or if their experiment allowed them to conclude something else was going on. The mathematics of probability (including permutations and combinations) had advanced to a point where it could be used to help answer such questions. This blend of necessity and available mathematics sparked the formulation of the field of statistics.

By its nature, statistics can be murky and ill defined. Broadly speaking, statistics uses mathematics to generate numerical measures that allow for decision making. These measures are constructed or conjectured by statisticians and are then tested to determine if they accurately measure what they were designed to measure. These measures may be refined to better reflect what the statistician had in mind.

Sometimes this constructive nature of the discipline causes people to be skeptical of statisticians, or believe they are deliberately manipulating data in order to fool the general public. Although statistics can be used in a misleading way, statisticians themselves are not trying to trick anyone. Assembling large amounts of data and describing them in a more useful, portable way is a difficult task, and statisticians haven't been doing it for that long. Statisticians have made a remarkable amount of progress, given the amount of time spent thus far. Two hundred years may seem like a long time to us, but compared to the five thousand or so years for which mathematics has been developing, statistics is very much the newcomer to the party.

Despite its youth, statistics has already grown enough that to discuss the entire field would take much more space than we have in this section of the *Mathematics Resource Guide*. As usual, our goal here is not to have the reader learn all there is to know about statistics, but to begin from a place typically studied in high school mathematics and end a bit beyond high school mathematics, with the hopes of making an undergraduate course in statistics more accessible to the reader.

3.1 DESCRIPTIVE STATISTICS

One aspect of statistics is an effort to communicate quickly and effectively about potentially large sets of data. We need to do something to describe a large data set to someone besides showing him or her all 1000 pieces of data and saying, "Look!" It seems there are two important ideas about the data set we would like to communicate. The first idea is what a typical value in this data set is, and the second is how spread out the values in this data set are. Together, these measures attempt to describe a set of data without listing every data point—hence the name descriptive statistics.

3.1.1 MEAN, MEDIAN, AND MODE

We will begin with the measures most commonly introduced in high school (or even middle school) mathematics: mean, median, and mode. Although we assume the reader has some familiarity with these measures, the rationale behind these measures and an understanding of why all three are used is not as commonly studied in high school mathematics. We hope the reader will gain an appreciation of all three measures and a deeper understanding of why multiple measures of a "typical" data value are necessary.

Let's consider an example. A small company has twenty employees. The annual salary of each employee is listed below in thousands of dollars, sorted from highest to lowest. What does a typical employee of this company earn per year?

SALARIES (IN THOUSANDS OF DOLLARS)				
\$25	\$35	\$40	\$50	\$58
\$28	\$36	\$42	\$53	\$62
\$31	\$36	\$44	\$54	\$65
\$32	\$36	\$45	\$55	\$173

How do we describe what a typical employee in this company earns? This depends, of course, on our definition of typical. There are several ways to answer this question. One way is to pretend that every employee in the company makes the same amount of money. To do this, we would find the sum of all the salaries and then divide by the number of employees. Finding the sum of all the employees is easy enough to do, but notionally we would like to do something besides write out 25 + 28 + 31 + … + 65 + 175. The similarity to a series is hopefully clear: if we pretend the salaries are a sequence, where x_1 is the salary of the first employee (\$25,000), x_2 is the salary of the second employee (\$28,000), and so on, we can quickly represent the series as $\sum_{i=1}^{20} x_i$ which equals \$1,000,000. If every employee made the same amount of money, this \$1,000,000 would be evenly distributed over all 20 salaries, and therefore each employee would make \$50,000. This value is referred to as the **mean** of the data set.

DEFINITION AND FORMULA FOR THE MEAN OF A DATA SET

The **mean** of a data set is the value each data point would have if all data points were equal. To find the mean, we sum the data values and then divide by the number of data points in our set.

Therefore, the mean of a set consisting of n values is given by $\frac{\sum_{i=1}^{n} x_i}{n}$, where x_i represents the i^{th} data value. The mean of a set of data is represented by $\bar{x}$ or sometimes the Greek letter μ (pronounced "mu").

Why might the mean not be a good representation of a "typical" data value? Comparing the mean salary of $50,000 to the actual distribution of salaries given in our example should make us question whether a typical employee makes $50,000. Although the mean salary is $50,000, the majority of the employees make less than this amount. Even if we think about the mean as the "center" of the data, this set of salaries is not centered on $50,000. For example, if we consider the values within $10,000 of $50,000, there are an equal number of values between $40,000 and $50,000 as there are between $50,000 and $60,000 within this interval, but far more values lie below $40,000 (8) than above $60,000 (3).

What is it about these data that causes the mean to be a potentially inaccurate measure of a typical data value? We note that the largest salary, $173,000, is significantly larger than all other data values in the set. This type of data point, called an **outlier**, is a value that is significantly different from the rest of the data set. There is some discussion about how to precisely determine what is considered an outlier, and we will discuss the most generally accepted method later on in this section. Under almost any definition or mathematical procedure to determine an outlier, the salary of $173,000, being almost three times the next closest salary, is considered an outlier.

Why does the outlier impact the mean? Since the mean is based on the sum of all of the data, an outlier will contribute much more to this total than a data point closer to the overall group. The larger the data set, the smaller this impact will be (since the sum is divided by the number of data points), but this can still significantly alter the mean of the data. For example, if the $173,000 salary is removed from the data above, the mean becomes $43,526. This new mean has almost the same number of values above it as below, and it is more centered within the data.

This idea of being centered in the data leads to another idea for measuring a typical data value. Reporting the middle data value allows for a measure that is not affected by the presence of outliers. Because the middle number does not depend on the values of the data points, but only how many data points there are, it does not matter if the highest salary is $65,000 or $173,000 or even $1,000,000. This measure of a typical data value is called the **median**. If the data set contains an odd number of values, the median is the value in the middle of the data. If the data set contains an even number of values, the median is the mean of the two values in the middle.

DEFINITION AND FORMULA FOR THE MEDIAN OF A DATA SET

The **median** of a set of data is the value with the same number of data points above as below the value. If the data set has n values and n is odd, then the median is $x_{\frac{n+1}{2}}$. If the data set has n values and n is even, the median is the mean of the two data values at the middle of the data: $\dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$.

In order to properly find the median of a data set, the data must be organized from smallest to largest value. The median of the salary data given in our example is $43,000.

The median is often used together with the mean to give information about the distribution of values in the set of data. Since the median is always in the exact middle of the data set, if the mean is close to the median, then the data are fairly evenly distributed on both sides of the median. If the mean is higher than the median, then the data are skewed above the median because the mean is being increased by values larger than the median. If the mean is less than the median, then the data are skewed below the median, as the mean is being decreased by values smaller than the median.

There is one more possible interpretation of what it means to be a typical data value: the data value that occurs most often. Although this value is not generally reported alone as the only representation of the typical data value, it is a typical value in the sense that if a value is selected at random, it is the value most likely to be selected. This value is called the **mode**.

DEFINITION OF MODE

The **mode** of a data set is the value that occurs most frequently

If each value in the data set occurs once, the data have no mode. If two different data values each occur most frequently, they are both considered the mode, and the data set is called **bimodal**. The mode of the salary data in our example is $36,000.

The mean, median, and mode are the three measures of central tendency generally used in statistics. They all attempt to capture and report what a typical data value is for the data in question. Each of these measures reports something slightly different about the data, and so they are often reported together to give a more complete picture of the data.

EXAMPLE 3.1A: Jill would like to average 85 out of 100 on her five science tests. She has already taken the first three tests and scored 89, 91, and 78. What does Jill need to average on her last two tests in order to have an overall average of 85?

SOLUTION: In order to average an 85 over five tests, Jill needs to earn a total of $85 \cdot 5 = 425$. Thus far she has earned a total of $89 + 91 + 78 = 258$, so she needs to total 167 on her last two tests. This means she needs to average $167 / 2 =$ **83.5** on her last two tests in order to have an overall average of 85.

EXAMPLE 3.1B: The median of the data set consisting of 3, 14, 20, 52, and x is x. What is the value of x?

SOLUTION: The best way to approach this problem is to think about the possible ways the data could be arranged. In order to find the median, the data need to be arranged in order from lowest to highest value. Therefore, if x is less than 3, the data would read: x, 3, 14, 20, 52.

In this case, the median would be 14, but x was assumed to be less than 3 in this scenario, so this situation is impossible. A similar argument shows that if x was larger than 52, the median would be 20, but since x cannot simultaneously be 20 and greater than 52, this is also impossible.

What if x were between 3 and 52? In that case, the data could look like one of the three following scenarios:

Scenario I: 3, x, 14, 20, 52

Scenario II : 3, 14, x, 20, 52

Scenario II : 3, 14, 20, x, 52

Under Scenario I, the median of the data set is 14. Therefore, if $x = 14$, the median would equal the value of x. Under Scenario II , the median of the data set is x. Therefore, if x is any number between 14 and 20, the median would equal the value of x. Under Scenario III , the median of the data set is 20. Therefore, if $x = 20$, the median would equal the value of x. So, x could be **14, 15, 16, 17, 18, 19,** or **20.**

Example 3.1c: Describe all possible data sets with more than one data point that have the same value for the mean, median, and mode.

Solution: Before describing all such possible sets, first we need to convince ourselves that such a set exists. Once we believe this is possible, then we can think up a few more and start to describe what all such sets look like.

Why must the data set have more than one data point? Let's suppose our data set consists of a single value; say, 4. The mean of this data set is 4, the median is 4, and the mode is 4. So this data set would have the same mean, median, and mode. Certainly there is nothing special about the choice of 4, so any data set with a single data point will have the same mean, median, and mode. This is sort of interesting, but not very realistic. The main purpose of descriptive statistics is to quickly communicate information about a data set instead of listing all the data points. If there is only one data point, communicating the entirety of the data isn't too difficult.

So, what if our set had two data points? In this case, the mean and the median are always the same since by definition the median of a data set with an even number of values is the mean of the middle two numbers. However, if the two data points were different, there would be no mode. Therefore, in order for the mean, median, and mode to be the same, both data points must be the same value. That isn't all that interesting, but neither is a set with only two data points.

What if our data set had three data points? The median of three values will automatically be the middle value. How can the mean equal this middle value? The mean is the sum of all the values divided by three, so the sum of the values must be three times the median. For this to occur, all three values need to be the same, or the other two values must differ from the median by the same amount. This way, the total sum will still equal three times the median. For example, the data set 21, 25, 29 has a median of 25 and a sum of 75, which gives us a mean of 25.

So, if the three data values are evenly spaced about the center value, the mean will equal the median. But, this implies the three data values are different, so therefore there is no mode! This means the only way to make a data set with three values with mean, median, and mode equal is to have three identical data values. Again, this is not a very interesting data set.

But, with four data points something interesting starts to occur. The set 21, 25, 29 has a mean and a median of 25, so if we want to make the mode 25, we can add another value of 25 to the data set. The set 21, 25, 25, 29 has mean, median, and mode equal to 25. Can we continue adding values but maintain a mean, median, and mode of 25? Certainly we can add as many 25's as we want, but this will become boring after a while. Let's think about the possibilities for 9 data points. Other than a data set consisting of nine values of 25, what else can we construct with mean, median, and mode of 25?

In order to have a median of 25 with nine values, the middle value must be 25. We also want the mode to be 25, so we need to have at least one more 25. With two values of 25, however, there will be an unequal number of data values above the median as below, say four values greater than 25 and three less than 25. Although we can pick values such that the mean will still be 25 with this imbalance, it is much easier if there are the same number of values above the median as below. Therefore, we would prefer three values of 25, so we can have three values below 25 and three above.

To have a mean of 25 requires that the data sum to $9 \cdot 25 = 225$, and with three values already fixed at 25, this means the remaining values must sum to 150. The three values larger than 25 can sum to whatever value we wish, as they can be as large as possible, but the three values less than 25 must sum to at least 1. (If all three values were 0, then 25 would not be the only mode.) So again we are faced with a large number of possibilities, but if we select nice possibilities, something interesting begins to happen.

Based on our smaller example with four data values, we start with the data set 21, 23, 25, 25, 25, 27, 29. This has a median and mode of 25, and a quick check verifies that the mean is also 25. This data set only has seven values, however, so we need to add two more data values. If both of these values were placed on the same side of 25, either both above or both below, this would cause the mean to move away from 25. Therefore, these two additional data points must be placed evenly around 25, as was done when there were only three data points. As long as these two data values sum to 50, they can be placed anywhere, but it seems nicer if they are placed with our original data, as in: 21, 23, 23, 25, 25, 25, 27, 27, 29.

So, although we may lack technical statistical knowledge to specifically describe data sets where the mean, median, and mode are all equal, we can say that these types of data sets seem to have some form of symmetry. In order to have equal mean and median, values need to be evenly spaced about the median, and for the mode to be the same value, there needs to be a spike at the center of the data.

Although there are several different important statistical distribution patterns, the Normal Distribution has this general shape and the important characteristic that the mean, median, and mode are all equal. We will study the Normal Distribution at the end of this resource guide.

Our experience with constructing data sets where the mean is equal to the median has introduced us to another interesting idea: what is the mean and median of a finite arithmetic sequence? Based on our preliminary work in the previous example, it seems that the mean and median of an arithmetic sequence will be equal since the values will be evenly distributed above and below the mean. Let's look at one more example to see if this idea is correct.

Example 3.1d: Consider the following data: 3, 7, 11, 15, 19, 23, 27, 31, 35. What are the mean and median of these data?

Solution: As these values are already arranged in increasing order, it is straightforward to read the median as **19**. Is the mean of these data also 19? We know from prior work in this resource guide that to find an arithmetic series, we create pairs with a constant sum. In this case, the paired sum is 38, so the sum of the terms will be $\frac{38 \cdot 9}{2}$, which is $19 \cdot 9$. Since there are 9 terms, we divide by 9 to find the mean, and the mean is indeed **19** as we wanted.

Is this always true? A generic arithmetic sequence with n terms, first term of a and constant difference of d will sum to $\frac{[a + a + (n-1) \cdot d] \cdot n}{2}$, which means the average will be $\frac{[a + a + (n-1) \cdot d] \cdot n}{2n}$.

This simplifies to $\frac{[2a + (n-1) \cdot d]}{2}$ or $a + \frac{(n-1)}{2} \cdot d$.

We would like this value to be the median of our arithmetic sequence as well, so it needs to be positioned in exactly the middle of the data. We recall that the direct formula for an arithmetic sequence is $x_k = a + (k-1) \cdot d$ since to get from the first term to the k^{th} term requires adding the difference $k-1$ times. The index for our arithmetic sequence is running from 1 to n, which means the index for the median needs to be $\frac{n+1}{2}$. (Think through a few examples to convince yourself of this!) But this is the index, not the value of the term in our sequence. So what is $x_{\frac{n+1}{2}} = ?$ Substituting $\frac{n+1}{2}$ for k in $x_k = a + (k-1) \cdot d$ yields $x_{\frac{n+1}{2}} = a + \left(\frac{n+1}{2} - 1\right) \cdot d$, which simplifies to $a + \left(\frac{n-1}{2}\right) \cdot d$, or $a + \frac{(n-1)}{2} \cdot d$ as desired. Therefore, the mean and the median of a finite arithmetic sequence are always equal.

This can make it somewhat easier to answer our previous question about data sets where the mean, median, and mode are all equal. Provided we begin with an arithmetic sequence, we can then add extra data points matching the median, so the mode is the same as the mean and the median. Additional values can then be added without disturbing the mean or median as long as they are added symmetrically about the desired median.

The mean, median, and mode all give information about a typical or representative data value and attempt to communicate information about the center of the data. In describing a data set, we are also interested in how the data are spread out. Therefore, we will now turn our attention to measures that try to communicate information about the distribution of the data.

3.1.2 RANGE, QUARTILES, AND IQR

As our extended thought experiment regarding data sets with equal mean, median, and mode hopefully showed us, data sets that are very different may have the same mean, median, and mode. For example, consider the two data sets below:

Set 1: 0, 0, 1, 3, 8, 9, 12, 18, 20, 20, 20, 22, 26, 27, 30, 31, 32, 42, 59

Set 2: 15, 15, 16, 17, 17, 18, 18, 19, 20, 20, 20, 21, 22, 22, 23, 23, 24, 24, 26

Both of these data sets have the same number of values (19), and the mean, median, and mode of each data set is 20. Reporting only these values to describe the data sets therefore gives the impression that these two sets are very similar. Clearly these data sets are drastically different, but how can we describe this difference and how do we capture this difference mathematically?

What makes these data sets different from each other is the spread of the values: the first data set is very spread out, while the second data set is clumped fairly close together. In the next section, we will develop a single value that attempts to describe the amount of spread present in a set of data. First, however, we will approach this problem in a more straightforward manner.

The simplest measurement of how the data are spread out is to report the difference between the highest and lowest data value. For Set 1, there is a difference of 59 between the highest and lowest data value; for Set 2, this difference is only 11. This measure of the difference between the highest data value and lowest data value is called the **range**.

DEFINITION

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

The range gives a very rough guideline for the spread of the data. A small range means the data are grouped relatively close together, and a large range means the data could be spread out. Using the range as the only measure of the spread of the data has several problems, however. The first is that the range is easily affected by the presence of outliers since it finds the difference between the most extreme values, and these data points could be outliers. This might give the impression that the data are very spread out when in fact the majority of them are not.

For instance, the range of the following data: 14, 15, 15, 15, 16, 17, 18, 18, 19, 20, 20, 52 is 38, which gives the appearance of a disparate and spread out data set. As can be seen from the values, however, the outlier of 52 is masking the true spread of the data. This sensitivity to outliers is one weakness of the range, as it will occasionally over-report the spread of the data.

The other potential issue of the range is that in order to be properly interpreted, it needs to be reported with the mean. Since the range measures the raw difference between the maximum and minimum data value, this difference can indicate a larger or smaller spread depending on the values of the data. A range of 50 with a mean of 1000 looks very different from a range of 50 with a mean of 25. (Is it possible to have a data set with a mean of 25 and a range of 50?) Therefore, we turn our attention to another measure of the spread of the data: **quartiles**.

Let's consider again the two data sets from our earlier discussion:

Set 1: 0, 0, 1, 3, 8, 9, 12, 18, 20, 20, 20, 22, 26, 27, 30, 31, 32, 42, 59

Set 2: 15, 15, 16, 17, 17, 18, 18, 19, 20, 20, 20, 21, 22, 22, 23, 23, 24, 24, 26

What makes data Set 1 and Set 2 different from each other is the spread of the data. Although the median of each data set is 20, the values in Set 1 are much more spread out than the values in Set 2. If we picture the median dividing each data set in half, the values in the lower half of Set 1 are generally much less than the values in the lower half of Set 2. Similarly, the values above the median in Set 1 are generally much higher than the values above the median in Set 2. To capture this difference, we could report the median of the values below the median for Set 1 and Set 2. This could show that although the medians are the same, the values below the median are lower in Set 1 than in Set 2, demonstrating a greater spread of data. The median of the values above the median could also be reported for both sets, and again we would anticipate a higher median from the values above the median for Set 1 than Set 2. Let's see how this would work with the actual data.

The values below the median for Set 1 are: 0, 0, 1, 3, 8, 9, 12, 18, 20. The 20 is included here because the middle 20 in Set 1 is the median. The median of these values is 8.

The values below the median for Set 2 are: 15, 15, 16, 17, 17, 18, 18, 19, 20. The median of these values is 17.

Calling this measure "the median of the values below the median" is awkward, so for now we will call this the lower median. The fact that the medians of Set 1 and Set 2 are equal, but the lower median for Set 1 is 8 while the lower median for Set 2 is 17 tells us that the values in Set 2 are much closer to the median than Set 1. At least, it tells us this is true below the median. What about above the median?

The values above the median for Set 1 are: 20, 22, 26, 27, 30, 31, 32, 42, 59. The median of these values is 30.

The values above the median for Set 2 are: 20, 21, 22, 22, 23, 23, 24, 24, 26. The median of these values is 23.

As we predicted, the fact that the values in Set 1 are much more spread out than the values in Set 2 is reflected in the so-called upper median as well. An upper median of 30 for Set 1 compared to an upper median of 23 for Set 2 tells us that the values in Set 1 are much more spread out than the values in Set 2.

What do these three numbers (the lower median, the median, and the upper median) tell us about the data? For Set 1, these numbers are 8, 20, 30. The median finds the middle value in the data, so this means 50% of the data are below 20, and 50% of the data are above 20. Since the lower median is the median of the values below the median, this means 25% of the original data are below 8, and 25% are between 8 and 20. Similarly, 25% of the data are between 20 and 30, and 25% of the data are above 30. Therefore, these values separate the data into quarters; hence their more common name: **quartiles**.

DEFINITION OF QUARTILES

The **lower quartile** or **first quartile** (abbreviated Q_1) is the median value of the data below the median in a set. The **upper quartile** or **third quartile** (abbreviated Q_3) is the median value of the data above the median in a set.

Together with the median, the lower and upper quartiles separate the data into four equal parts, meaning the same number of data values are in each area (less than the first quartile, between the first quartile and the median, between the median and the third quartile, and above the third quartile). This explains why the upper quartile is sometimes called the third quartile, as the median is the second quartile.

The difference between the lower quartile and upper quartile is referred to as the **interquartile range**, abbreviated **IQR**. The IQR measures the spread of the middle 50% of the data in the same way that the range measures the spread of the entire data, but as IQR is based on medians, it is not susceptible to outliers in the same way that the range is. Indeed, IQR is used in a common statistical test for outliers. Although there are more complicated procedures, the use of IQR to test for outliers remains very common due to its simplicity and ease of calculation.

IQR TEST FOR OUTLIERS

Data points that are further than 1.5 times the IQR above the upper quartile or below the lower quartile are considered outliers. Written as a formula, data point x_i is an outlier if $x_i > Q_3 + 1.5 \cdot IQR$ or $x_i < Q_1 - 1.5 \cdot IQR$.

Let's now use the IQR test for outliers to determine if there are any outliers in Set 1 or Set 2:

Set 1: 0, 0, 1, 3, 8, 9, 12, 18, 20, 20, 20, 22, 26, 27, 30, 31, 32, 42, 59

Set 2: 15, 15, 16, 17, 17, 18, 18, 19, 20, 20, 20, 21, 22, 22, 23, 23, 24, 24, 26

Having previously calculated the lower and upper quartiles for these data sets, we can quickly determine that the IQR for Set 1 is 30 – 8 = 22. Therefore, $1.5 \cdot (22) = 33$, and in order to be an outlier a data value would need to be larger than 63 ($Q_3 + 1.5 \cdot IQR = 30 + 33$) or less than –25 ($Q_1 - 1.5 \cdot IQR = 8 - 33$). None of our data values are less than –25, and the maximum, 59, turns out to not be quite high enough to be an outlier.

For Set 2, the IQR is 23 – 17 = 6. Therefore, $1.5 \cdot (6) = 9$, and in order to be an outlier a data value would need to be larger than 32 ($Q_3 + 1.5 \cdot IQR = 23 + 9$) or less than 8 ($Q_1 - 1.5 \cdot IQR = 17 - 9$). As would be expected for data with a small spread, no values are in danger of being classified as outliers.

As nice as this test is, it does beg a question: why 1.5 times the IQR? Why not twice the IQR, or just the IQR itself? The use of the IQR makes good sense, as an outlier should be far removed from a typical value, and the IQR measures the range of the middle 50% of the data. But why a multiplier of 1.5, rather than any other number? Statistical legend is that the originator of this test justified the selection of 1.5 by saying that a multiplier of 2 looked like too much, but 1 wasn't enough.

Although it is this type of apparent subjectivity that has at times caused the general public to mistrust statistics, the creation of this type of measure is at the heart of statistics and is what makes the discipline vibrant. Statisticians create new mathematical formulas or procedures to model uncertainty and, as in this case, represent and communicate information about data sets. If the measures do not successfully and reliably do what they were designed to do, they are replaced by newer and better measures. As technology and computing power have become readily available, the types of data sets statisticians have been able to analyze and the types of questions that can be asked are also changing. An example of this is the idea of capturing the amount of variation in a data set in a single measure. We will now turn our attention to just this task.

3.2 MEASURES OF VARIATION

Thus far we have looked at descriptive statistics as a way to report general information about the center and spread of a set of data. However, the measures of spread we have developed so far are not particularly good at communicating this information. As we have seen, range is very sensitive to outliers and may overstate the actual spread of the data. Quartiles give us a nice way to informally measure the spread, but mostly only when in relation to another set of data. It seems we need a more formal numerical measure of variation, similar to the measures of center. In this part of the resource guide we will develop such a measure.

3.2.1 VARIANCE

Let's return to data Set 1 and Set 2:

Set 1: 0, 0, 1, 3, 8, 9, 12, 18, 20, 20, 20, 22, 26, 27, 30, 31, 32, 42, 59

Set 2: 15, 15, 16, 17, 17, 18, 18, 19, 20, 20, 20, 21, 22, 22, 23, 23, 24, 24, 26

Whatever measure of spread we develop, clearly Set 1 should have a larger numerical value than Set 2. What is it that makes Set 1 have a much greater spread than Set 2? In Set 1, many more values are far away from the center of the data than in Set 2. So, an initial idea for measuring the variation in a set of data is to measure the total distance the data values are away from the center. But which measure of center are we using? Fortunately for us, we can postpone this question for a little while since the mean, median, and mode of these two sets of data are all 20. Let's look at the total difference away from 20 for each data set.

SET 1	DIFFERENCE FROM CENTER		SET 2	DIFFERENCE FROM CENTER
0	–20		15	–5
0	–20		15	–5
1	–19		16	–4
3	–17		17	–3
8	–12		17	–3
9	–11		18	–2
12	–8		18	–2
18	–2		19	–1
20	0		20	0
20	0		20	0
20	0		20	0
22	2		21	1
26	6		22	2
27	7		22	2
30	10		23	3
31	11		23	3
32	12		24	4
42	22		24	4
59	39		26	6
	0	**SUM**		0

This is perhaps an unexpected result: both data sets have a total difference from the center of 0. Is this always true?

It depends on which measure of center we are using. Certainly this does not need to be true for the mode since the mode could be anywhere in the set where there is a large cluster of one particular data value. It also is not necessarily true for the median since the median does not depend on the particular values in the data set, just the number of values. If we imagine a data set with a given median, we can increase the largest value in the data set without changing the value of the median. This will, however, increase the difference between the last value in the data set and the median, so the sum of the differences will increase. Therefore, this sum of differences cannot always be 0.

What about the mean? Increasing any value in the data set will increase the value of the mean, so this seems potentially promising. Let's consider the case above for Set 1 in a slightly generalized version and see what happens.

First, each data value is subtracted from the mean: $x_1 - \bar{x}$, $x_2 - \bar{x}$, $x_3 - \bar{x}$, and so on until $x_{19} - \bar{x}$. Then these differences are all added together: $\sum_{i=1}^{19} (x_i - \bar{x})$. In other words, $x_1 - \bar{x} + x_2 - \bar{x} + x_3 - \bar{x} + \ldots + x_{19} - \bar{x}$. We can reorder this calculation as $x_1 + x_2 + x_3 + \ldots + x_{19} - \bar{x} - \bar{x} - \bar{x} - \ldots \bar{x}$. Since there are 19 terms in the original series, there are 19 copies of $\bar{x}$, so this can be rewritten as $\sum_{i=1}^{19} x_i - 19 \cdot \bar{x}$. But there are 19 terms in the list, so in order to find the mean, we found the sum of all the data values and divided by 19! Since $\bar{x} = \frac{\sum_{i=1}^{19} x_i}{19}$, we now have $\sum_{i=1}^{19} x_i - 19 \cdot \frac{\sum_{i=1}^{19} x_i}{19} = \sum_{i=1}^{19} x_i - \sum_{i=1}^{19} x_i$, so this indeed will always equal 0.

Our work to determine this did not rely on the data values, so this could just as easily refer to Set 2 as to Set 1. And there is certainly nothing special about these data sets having 19 values—our work could be rewritten with n in the place of 19. This would show that the sum of the differences between the mean and the data values is always 0, regardless of the data values themselves or the size of the data set.

SUM OF THE DIFFERENCES FROM THE MEAN

The sum of the differences between the data values and the mean of the data set is always 0. In other words, if the data set has n values, $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$.

Proof: $\sum_{i=1}^{n} (x_i - \bar{x})$ means subtract each data value from the mean and then add all of the differences together. Writing out a few terms in the summation to get an idea of what is going on yields: $\sum_{i=1}^{n} (x_i - \bar{x}) = (x_i - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \ldots + (x_n - \bar{x})$.

Notice there are n copies of $\bar{x}$. This sum can be rearranged as: $= x_1 + x_2 + x_3 + \ldots + x_n - n \cdot \bar{x}$.

The first part of this sum lends itself nicely to sigma notation, and as $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, this equals: $= \sum_{i=1}^{n} x_i - n \cdot \frac{\sum_{i=1}^{n} x_i}{n}$

The n's cancel, and therefore we are left with: $= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$.

All of this is very nice and interesting, but it doesn't help us solve our original problem: how to develop a numerical measure of the spread for a data set. We would like to use something with the difference between each value and the mean, but these differences sum to 0 for all sets of data. Can we fix this somehow?

In order to fix this, we need to remove the negative differences, or make the differences that were negative no longer negative. If every difference were positive, then the sum of these differences would not be 0.

Also, values that are very far away from the mean should be worth more than values that are very close to the mean. Two options come to mind: 1) take the absolute value of the differences prior to summing and 2) square the differences prior to summing. After much discussion, statisticians decided to square the differences rather than take the absolute value (although there is support among some statisticians for using the absolute value).

Let's take another look at our table, this time with the squares of the differences included.

SET 1	$x_i - \bar{x}$	$(x_i - \bar{x})^2$		SET 2	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
0	–20	400		15	–5	25
0	–20	400		15	–5	25
1	–19	361		16	–4	16
3	–17	289		17	–3	9
8	–12	144		17	–3	9
9	–11	121		18	–2	4
12	–8	64		18	–2	4
18	–2	4		19	–1	1
20	0	0		20	0	0
20	0	0		20	0	0
20	0	0		20	0	0
22	2	4		21	1	1
26	6	36		22	2	4
27	7	49		22	2	4
30	10	100		23	3	9
31	11	121		23	3	9
32	12	144		24	4	16
42	22	484		24	4	16
59	39	1521		26	6	36
388	**0**	**4242**	**SUM**	**388**	**0**	**188**

Now it seems we have a measure we can use! The sum of the squares of the difference between each data value and the mean is far greater for Set 1 than for Set 2, telling us that Set 1 is much more spread out than Set 2.

Notice that this measure controls for the mean, since the mean is subtracted from each value, allowing us to compare the variation of data sets that do not have the same mean. In this case, Set 1 and Set 2 had the same mean. But, let's say every data point in Set 2 was increased by 10. The means of Set 1 and Set 2 would no longer be the same, but we would still want to say that Set 1 was more spread out than Set 2.

Increasing every data value in Set 2 by 10 will also increase the mean by 10, which means the sum of the squares of the differences between the data values and the mean will be the same 188 we had earlier. So this measure of variation is independent of the mean of our data set.

In its current form, however, this measure requires both data sets to have the same number of values. One can imagine a collection of data that is as spread out as Set 2 but has a larger sum of squares of differences than Set 1 simply by virtue of having more data points. For example, if we add enough values between 15 and 26 to Set 2 in such a way as to not change the mean, we can generate a sum of squares of differences that is larger than 4,242. This should not imply, however, that Set 1 is less variable than Set 2, so we need to control for the size of the data set. In order to do this, we divide by the number of data points, creating the average of the squares of the difference between the data points and the mean. This statistical measure is called the **variance**.

DEFINITION AND FORMULA FOR VARIANCE

The variance of a set of data is the average of the squared difference between each data point and the mean. Variance is denoted by the Greek letter sigma squared, written as σ^2. In symbols,

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}.$$

For example, the variance of Set 1 is $\frac{4242}{19} = 223.263$. The variance of Set 2 is $\frac{188}{19} = 9.895$.

It does take time to become accustomed to variance values and interpreting how much a data set varies from this value. As we can see, Set 2 is very homogenous, so a variance of 9.895 is actually rather low. Set 1 is extremely varied, and it has a variance of over 200!

When calculating variance by hand (which should be done rarely and only when absolutely necessary), using the above formula is rather lengthy and time-consuming. Furthermore, it uses the mean in calculation, and if the mean is not a whole number (like 20), the mean will be rounded. This makes the calculations much worse and also introduces an element of error into the variance calculation. Therefore, we will develop an equivalent formula to make calculating the variance by hand slightly easier and less prone to error.

Let's focus on the numerator of the variance calculation: $\sum_{i=1}^{n}(x_i - \bar{x})^2$. Each data value is subtracted from the mean and then squared, and all these values are summed. In symbols:

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2 .$$

Expanding each of these binomials gives us:

$$= x_1^2 - 2x_1\bar{x} + \bar{x}^2 + x_2^2 - 2x_2\bar{x} + \bar{x}^2 + x_3^2 - 2x_3\bar{x} + \bar{x}^2 + \ldots + x_n^2 - 2x_n\bar{x} + \bar{x}^2 .$$

Rearranging terms a bit and recognizing that there are n copies of $\bar{x}^2$ yields:

$$= x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2 + n \cdot \bar{x}^2 - 2x_1\bar{x} - 2x_2\bar{x} - 2x_3\bar{x} - \ldots - 2x_n\bar{x}.$$

The first n terms in this arrangement can be collapsed nicely using sigma notation, and all of the remaining terms share a common factor of $\bar{x}$, so: $= \sum_{i=1}^{n} x_i^2 + \bar{x} \cdot (n \cdot \bar{x} - 2x_1 - 2x_2 - 2x_3 - \ldots - 2x_n)$.

Part of this looks familiar from our proof that $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$, but we have two copies of each x_i, not just one.

Splitting each term apart therefore gives us:

$$= \sum_{i=1}^{n} x_i^2 + \bar{x} \cdot [(n \cdot \bar{x} - x_1 - x_2 - x_3 - \ldots - x_n) - x_1 - x_2 - x_3 - \ldots - x_n].$$

But, the inner parentheses sums to 0, since $n \cdot \bar{x} = n \cdot \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n$.

Therefore, this simplifies to: $= \sum_{i=1}^{n} x_i^2 + \bar{x}(-x_1 - x_2 - x_3 - \ldots - x_n)$.

Factoring out the negative yields: $= \sum_{i=1}^{n} x_i^2 - \bar{x}(x_1 + x_2 + x_3 + \ldots + x_n)$.

And the parentheses lend themselves nicely to sigma notation: $= \sum_{i=1}^{n} x_i^2 - \bar{x} \cdot \sum_{i=1}^{n} x_i$

This is potentially a much nicer calculation to do by hand. Notice the first summation is the sum of the *squares* of the data values, while the second summation is the sum of the data values. Although this formula looks very nice, part of the reason to simplify or alter the original equation was to avoid using the mean.

Therefore, we can substitute $\frac{\sum_{i=1}^{n} x_i}{n}$ for $\bar{x}$ and get: $= \sum_{i=1}^{n} x_i^2 - \frac{\sum_{i=1}^{n} x_i}{n} \cdot \sum_{i=1}^{n} x_i$, or $\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$.

ALTERNATIVE FORMULAS FOR CALCULATING VARIANCE

The following calculations are equivalent to $\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$ for finding the variance of a data set:

$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^{\,2} - \bar{x} \cdot \sum_{i=1}^{n} x_i}{n} \text{ or } \sigma^2 = \frac{\sum_{i=1}^{n} x_i^{\,2} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n}.$$

Note that in the final version of this formula, the two sigmas are asking for very different calculations. $\sum_{i=1}^{n} x_i^{\,2}$ means to square each data value and then sum, whereas $\left(\sum_{i=1}^{n} x_i\right)^2$ asks for the sum of each data value, and then that sum is squared.

These three formulas are all equivalent, but it should be stressed that the variance of a set of data should only be calculated by hand when completely necessary. Prior to the advent of computing technology, a need to accurately calculate variances by hand generated an interest in these alternative formulas. Almost all hand-held calculators and many computer programs have the ability to calculate the variance (and mean, median, mode, and quartiles). The reader should take time to become familiar with whatever calculator technology is available and make sure these statistical measures can be properly found using technology.

3.2.2 STANDARD DEVIATION

Although variance is a robust measure of the variation in a data set, it is a bit difficult to properly interpret what the variance means in the context of the data. This is partially because, as we saw earlier with Set 1, sometimes the variance is an extremely large number relative to the values in the data set. Saying Set 1 has a variance of 223.263 is nice enough when we are comparing it to the variance for Set 2, but when the data values in Set 1 range from 0 to 59, saying the variance is 223.362 doesn't really tell us anything useful about the values in Set 1.

The value of the variance can be so large because the differences between the data values and the mean are squared, which often creates terms much larger than any of the values in the data set. As we can see in the table for Set 1, the data value of 59 contributed 1521 to the sum of the squares of the differences. Even dividing this contribution by 19 to control for the size of the data set still results in 80.05263, a value larger than 59, the largest value in the data set. And 59 wasn't even an outlier for Set 1!

So, we would like some way to make the variance value smaller in order to better relate it to the original values in the data set. As we squared data values (well, technically differences of data values), it seems that taking the square root of the variance would be possible to justify. Let's see if we can figure out what the units in our variance calculation are and see if finding the square root of the variance returns our units to the same as those in the original data.

Assume Set 2 is the height, in inches, of a series of plants. Each data value is therefore in terms of inches: 15 inches, 15 inches, 16 inches, and so on. The mean of this data set will be 20 inches, and so the difference between each data value and the mean will also be in terms of inches, although perhaps "negative inches," which is a bit strange. But, we squared these differences in order to avoid this negative problem, so we end up with square inches as our unit: 25 inches squared, 25 inches squared, 16 inches squared, and so on. Finding the mean does not change the units, so the variance of Set 2 is 9.895 inches squared. Taking the square root of the variance, then, allows us to return to the units of the original data. The square root of the variance is called the **standard deviation**.

DEFINITION OF THE STANDARD DEVIATION

The standard deviation of a data set, denoted by the Greek letter sigma, is the square root of the variance for the data. In symbols, this is $\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$

The standard deviation of Set 1 is $\sqrt{223.263} = 14.942$. The standard deviation of Set 2 is $\sqrt{9.895} = 3.146$.

If necessary, the range can be used to quickly generate an extremely rough approximation of the standard deviation for a data set. In general, the standard deviation is approximately one-quarter of the range. This is an extremely rough approximation and should only be used if calculating the exact standard deviation is impossible or not necessary.

USING THE RANGE TO APPROXIMATE THE STANDARD DEVIATION

A crude approximation of the standard deviation is the range divided by 4; $\sigma = \dfrac{range}{4}$.

The range of Set 1 is 59, and $\frac{59}{4} = 14.75 \approx 14.942$. The range of Set 2 is 11, and $\frac{11}{4} = 3.75 \approx 3.146$. As can be seen from these two examples, sometimes this approximation is fairly good, as with Set 1 (about 1% error), but other times this approximation is not so good, as with Set 2 (about 20% error). As such, it should be used sparingly and with caution.

Like the variance, the standard deviation is used to quantify the consistency of a data set. A low standard deviation indicates the data are clustered close together while a higher standard deviation is caused by data that are more spread out. In cases where consistency is desired, then, the standard deviation takes on increasing importance.

Let's consider an example. Two companies, Crazy Cam's and Steady Sam's, manufacture PVC pipe that is 3 inches in diameter. Clearly there is some variation in PVC pipe manufacturing, as not every pipe will be completely identical. Ten pipes are selected at random from the most recent batch at each company, and the diameters of the pipes are measured. The data are shown below.

Crazy Cam: 1, 1, 1, 1, 1, 5, 5, 5, 5, 5

Steady Sam: 2.9, 2.9, 2.9, 2.9, 2.9, 3.1, 3.1, 3.1, 3.1, 3.1

Both of these data sets have a mean (and median) of the specified 3 inches, but clearly those purchasing PVC pipe will select Steady Sam's rather than Crazy Cam's. The standard deviation of Crazy Cam's data is 2 (think about why this is true without actually calculating it), while the standard deviation of Steady Sam's data is 0.1 (again, think this through without calculating).

Although low standard deviations are desirable when seeking consistency, extremely low standard deviations (like 0.1) are, in fact, quite rare in any form of naturally occurring data. In order for the standard deviation to be 0, all data values would need to be identical, and even several small variations away from the mean start to increase the standard deviation.

The standard deviation also can act as a way to compare the relative position of two data values relative to their data sets, as we will discuss next.

3.2.3 Z-SCORE

Taylor took a math test and a science test. Let's try to determine which test Taylor performed better on. Taylor scored 58 on the math test and 56 on the science test. On which test did Taylor perform better? In the absence of more information, it appears Taylor did better on the math test.

Let's take another look at this problem, but with different information. Taylor scored a 58 on the math test, and the average score on the math test was 54. Taylor scored 56 on the science test, and the average score on the science test was 50. On which test did Taylor perform better? Now it seems that Taylor did better on the science test, since the score on the science test was 6 points above the mean, while the score on the math test was only 4 points above the mean.

Now, let's consider this with a bit more information: Taylor scored a 58 on the math test. The average score on the math test was 54, and the standard deviation of the scores was 1.5. Taylor scored 56 points on the science test. The average score on the science test was 50, and the standard deviation of the scores was 5. On which test did Taylor perform better? Although Taylor's score on the math test is closer to the mean than the science score, we also know that the math scores were clustered much closely together than the science scores. On the math test, Taylor performed better than two standard deviations above the mean, while on the science test Taylor performed a bit better than one standard deviation above the mean. Therefore, Taylor did better on the math test than the science test.

Historically, this is what "grading on the curve" meant. Although used inappropriately now to mean something like adding points to every score, to grade on a curve really means to assign grades based on how students do relative to each other. Usually, scores that were two standard deviations above the mean received a grade of 'A', scores that were between one and two standard deviations above the mean received a grade of 'B', scores that were within one standard deviation of the mean were 'C's, scores that were between one and two standard deviations below the mean were 'D's, and scores worse than two standard deviations below the mean were assigned 'F's.

The idea that Taylor scored about two standard deviations above the mean on the math test and about one standard deviation above the mean on the science test is called a **z-score.**

DEFINITION AND FORMULA FOR Z-SCORE

A **z-score** represents the number of standard deviations a particular data value is above or below the mean. A z-score is calculated by the formula $z = \frac{x_i - \overline{x}}{\sigma}$. A z-score is also sometimes called a **standard score**.

Taylor's z-score on the math test is $\frac{58 - 54}{1.5} = 2.667$. Taylor's z-score on the science test is $\frac{56 - 50}{5} = 1.2$.

We can notice a few items about z-scores right away. Z-scores equal 0 if the data value is the same as the mean, and z-scores are positive or negative if the data value is above or below the mean, respectively. Z-scores will almost always be non-whole values, based on the division involved. Means and standard deviations themselves are already rarely whole numbers, and a division of two non-whole numbers will, more often than not, produce another non-whole number. Therefore, z-scores are not something to try to calculate by hand unless absolutely necessary. Furthermore, most z-scores are relatively small, between –1 and +1. It is somewhat difficult for z-scores to be greater than 1 or less than –1, difficult for z-scores to be greater than 2 or less than –2, and extremely difficult for z-scores to be greater than 3 or less than –3. To illustrate this, let's look at the data from Set 1 and Set 2 again with the corresponding z-scores for each data value.

SET 1	Z-SCORE		SET 2	Z-SCORE
0	**–1.339**		15	**–1.590**
0	**–1.339**		15	**–1.590**
1	**–1.272**		16	**–1.272**
3	**–1.138**		17	**–0.954**
8	**–0.803**		17	**–0.954**
9	**–0.736**		18	**–0.636**
12	**–0.535**		18	**–0.636**
18	**–0.134**		19	**–0.318**
20	**0**		20	**0**
20	**0**		20	**0**
20	**0**		20	**0**
22	**0.134**		21	**0.318**
26	**0.402**		22	**0.636**
27	**0.468**		22	**0.636**
30	**0.669**		23	**0.954**
31	**0.736**		23	**0.954**
32	**0.803**		24	**1.272**
42	**1.472**		24	**1.272**
59	**2.610**		26	**1.907**
	0	**SUM**		**0**

Out of all these data points, only 59 from Set 1 has a z-score greater than 2. Indeed, this data value is so far removed from most of Set 1 that initially we were concerned the 59 was an outlier! In Set 2, however, something strange is happening. Although the data points are clustered close together, the z-scores for individual values are larger than we might perhaps expect. The value of 24 in Set 2, for instance, has a z-score of 1.272, while the value of 32 in Set 1 has a z-score of 0.803. The 32 in Set 1 is much farther away from the mean than the 24 in Set 2, so what is happening? The z-score measures how many standard deviations a value is away from the mean. So, the very fact that Set 1 is spread out is what allows for values to be further away from the mean; it is easier to be away from the mean in Set 1, so we should be less surprised by values that stray from the mean. This causes the z-score for the 32 to be somewhat lower than we might expect. By contrast, in Set 2, the data are clustered very close together, and the standard deviation is very low. Therefore, it is more difficult to stray very far from the mean, so the z-score for the 24 is higher than we might expect.

Notice also that the sum of the z-scores for both these data sets is 0. Hopefully this does not surprise us upon reflection. If we imagine the sum of all the z-scores for, say, Set 2, we have $z_1 + z_2 + z_3 + \ldots + z_{19}$. Each of these z-scores measures the number of standard deviations away from the mean each data value is, so $z_1 = \frac{x_1 - \bar{x}}{\sigma}$, $z_2 = \frac{x_2 - \bar{x}}{\sigma}$, and so on until $z_{19} = \frac{x_{19} - \bar{x}}{\sigma}$. To sum all these fractions together looks ugly until we realize they all have the same denominator, so: $\frac{x_1 - \bar{x}}{\sigma} + \frac{x_2 - \bar{x}}{\sigma} + \frac{x_3 - \bar{x}}{\sigma} + \ldots + \frac{x_{19} - \bar{x}}{\sigma} = \frac{x_1 + x_2 + x_3 + \ldots + x_{19} - n \cdot \bar{x}}{\sigma}$.

By now we have seen the numerator of this fraction more than once, and so we recognize that it sums to 0. Therefore, the sum of the z-scores is always 0 for any data set.

Example 3.2a: Assume the average height of people in a major metropolitan area is 68.5 inches, with a standard deviation of 2.5 inches. What is the z-score for an individual with a height of 72 inches?

Solution: $z = \frac{x - \bar{x}}{\sigma}$, so $\frac{72 - 68.5}{2.5} =$ **1.4.**

Example 3.2b: Assuming the information from the Example 3.2a holds, how tall would someone need to be in order to have a z-score of –3?

Solution: $z = \frac{x - \bar{x}}{\sigma}$, so $-3 = \frac{x - 68.5}{2.5}$. Solving for x yields **61 inches**.

It is natural to wonder at this point how likely it is to find someone in this population who is 61 inches tall, or, to phrase it more generally, how likely it is to have a z-score of –3. The relationship between z-scores and probabilities is complicated, but it is at the heart of statistics. We will by no means answer this question fully during the remainder of the *Mathematics Resource Guide*, but we hope to give the reader some background and sense of this important, fundamental connection in statistics. To answer this specific question about the probability of finding someone in this population who is 61 inches or taller, however, we first need some understanding of probability.

3.3 BASIC PROBABILITY

The chance that an event will or will not occur is a seemingly simple concept, but probability becomes fairly complicated quickly. In this section, we will explore introductory probability concepts in order to help us better understand the mathematics of chance. Once we are familiar with the rules of probability, we will look at probability distributions, a foundational concept in statistics.

At its core, probability consists of a single formula: the probability of an event occurring is the number of ways that event can occur divided by the total number of possible outcomes. In symbols, the probability of an event E occurring is given by $p(E)$, where $p(E) = \frac{\textit{number of ways event E can occur}}{\textit{total number of outcomes}}$. Although mathematically simple, this formula is capable of dealing with a large variety of probability problems. Counting the total number of outcomes or the number of ways the event E can occur can become more complicated, but our basic probability structure will remain the same. We will begin with a few examples.

EXAMPLE 3.3A: What is the probability of rolling a fair die numbered 1–6 and rolling a prime number (1 is not prime)?

SOLUTION: Since there are three prime numbers on a die numbered 1–6 (2, 3, and 5), the probability of rolling a prime number is $\frac{3}{6} = \mathbf{\frac{1}{2}}$.

EXAMPLE 3.3B: What is the probability of drawing a face card from a standard deck of playing cards?

SOLUTION: As there are 12 face cards in a standard deck of playing cards (Jack, Queen, and King of spades, clubs, diamonds, and hearts), the probability of drawing a face card is $\frac{12}{52} = \frac{3}{13}$, or about **0.2308**.

EXAMPLE 3.3C: A bag contains 12 blue marbles, 8 white marbles, and 5 yellow marbles. What is the probability of drawing a white marble from the bag?

SOLUTION: Since there are 8 white marbles out of 25 total marbles, the probability of drawing a white marble is $\frac{8}{25} =$ **0.32**.

EXAMPLE 3.3D: Using the same bag of marbles from EXAMPLE 3.3C, what is the probability of drawing a blue marble?

SOLUTION: There are 12 blue marbles and 25 total marbles, so the probability of drawing a blue marble is $\frac{12}{25}$, or **0.48**.

EXAMPLE 3.3E: Using the same bag of marbles, what is the probability of drawing a yellow marble?

SOLUTION: There are 5 yellow marbles, so the probability of drawing a yellow marble is $\frac{5}{25}$, or **0.2**.

Example 3.3f: Using the same bag of marbles, what is the probability of drawing a green marble?

Solution: As there are no green marbles in the bag (and we are not allowed to combine a blue marble and yellow marble), the probability of drawing a green marble is **0**.

Example 3.3g: Using the same bag of marbles, what is the probability of drawing a marble that is not blue?

Solution: Since there are 13 marbles that are not blue, the probability of drawing a non-blue marble is $\frac{13}{25}$ or **0.52**. Notice that $p(\textit{blue}) + p(\textit{not blue}) = 1$.

Example 3.3h: Using the same bag of marbles, what is the probability of drawing a blue, white, or yellow marble?

Solution: Since there are 25 marbles that are blue, white, or yellow, the probability of drawing a blue, white, or yellow marble is **1**.

These examples illustrate the following basic properties of probability:

BASIC PROPERTIES OF PROBABILITIES

The probability of any event is a number between 0 and 1. An event with probability 0 is impossible, and an event with probability 1 is certain. In symbols, for any event E, $0 \le p(E) \le 1$.

The probability of an event occurring plus the probability of the event not occurring is 1. The probability of an event E not occurring is called the **complement** of E and is denoted by E'. In symbols, $p(E) + p(E') = 1$.

In a given situation, the sum of the probabilities of all possible events must equal 1, since one of the possible events must occur. In symbols, if there are n different possible outcomes, $\sum_{i=1}^{n} p(E_i) = 1$.

These properties are clear if we consider a specific context where it is possible for us to imagine the set of all possible outcomes, like the bag of marbles. However, we will also deal with probability situations in which imagining or constructing the set of all possible outcomes is either impossible or extremely difficult. Therefore, we will develop properties like the ones just listed to help us deal with more complicated probability problems. Next we will develop a few rules of basic probability. Again, these rules will be developed from a context where they are not necessary, but once developed these rules can be applied to more general probability situations.

3.3.1 INDEPENDENT EVENTS

While we use our initial definition for probability, $p(E) = \frac{\text{number of ways event } E \text{ can occur}}{\text{total number of outcomes}}$ for the probability of a single event happening, how do we proceed when there are two (or more) events happening? When there are multiple events, there are two possibilities:

1. The occurrence (or non-occurrence) of one event does not affect the occurrence (or non-occurrence) of the other event(s).

2. The occurrence (or non-occurrence) of one event does affect the occurrence (or non-occurrence) of the other event(s).

These two situations are represented in the following examples.

EXAMPLE 3.3I: In a board game, two fair six-sided dice are rolled. The first die is numbered 1–6, and the second die has two faces each of the following three colors: green, red, and blue. What is the probability of rolling the two dice and getting an even red result?

EXAMPLE 3.3J: Two cards are drawn from a standard deck of 52 playing cards without replacement (meaning the first card is not put back in the deck before the second card is drawn). What is the probability of drawing two kings?

Both examples require two outcomes. In EXAMPLE 3.3I, the two outcomes are an even number and a red face. In EXAMPLE 3.3J, the two outcomes are drawing a king and then another king. However, in EXAMPLE 3.3I, the result on the numbered die does not affect the result on the colored die, or vice versa. In EXAMPLE 3.3J, drawing an ace on the first draw changes the probability of drawing an ace on the second draw. We call events like those in EXAMPLE 3.3I **independent**, and events like those in EXAMPLE 3.3J **dependent**.

DEFINITION OF INDEPENDENT EVENTS

Two (or more) events are called **independent** if the occurrence (or non-occurrence) of one event does not affect the occurrence (or non-occurrence) of the other event(s). When determining the probabilities of each event, the events can be considered separately since there is no meaningful interaction between the events.

DEFINITION OF DEPENDENT EVENTS

Two (or more) events are called **dependent** if the occurrence of one event does affect the occurrence of the other event(s). When determining the probabilities of each event, the events cannot be considered separately, as the events are connected in a meaningful way.

In the next section of the resource guide, we will examine the probabilities of dependent events. First we will focus on only independent events. Let's begin with Example 3.3i from earlier.

Example 3.3i (Revisited): In a board game, two fair six-sided dice are rolled. The first die is numbered 1–6, and the second die has two faces each of the following three colors: green, red, and blue. What is the probability of rolling the two dice and getting an even red result?

Solution: Using our definition of probability, $p(E) = \dfrac{\textit{number of ways event E can occur}}{\textit{total number of outcomes}}$, requires counting the number of possible ways to get an even red result and the total number of outcomes. Since both dice have six sides, the Multiplication Rule tells us there are $6 \cdot 6 = 36$ total outcomes. As there are three ways to get an even number on the numbered die and two ways to get a red result from the colored die, again the Multiplication Rule tells us there are $3 \cdot 2 = 6$ ways to get an even red result. Therefore, the probability of getting an even red result is $\frac{6}{36} = \mathbf{\frac{1}{6}}$.

Let's see if there is a relationship between the combined probability of getting an even red result and the individual probabilities of getting an even number on the numbered die and rolling a red face on the colored die. For the numbered die, three of the six faces are even numbers, so the probability of rolling an even number is $\frac{3}{6} = \frac{1}{2}$. For the colored die, two of the six faces are red, so the probability of rolling a red face is $\frac{2}{6} = \frac{1}{3}$. As $\frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$, it seems we should multiply the probabilities. Is this a coincidence, or does it make mathematical sense? Let's look at another example.

Example 3.3k: Each morning before going to work, Mr. Scott selects his outfit at random. He has nine shirts: four blue, three white, one yellow, and one green. He has seven pairs of pants: four khaki, two black, and one grey. He has two pairs of shoes: one brown and one black. What is the probability that Mr. Scott will wear a blue shirt with black pants and brown shoes?

Solution: Using $p(E) = \dfrac{\textit{number of ways event E can occur}}{\textit{total number of outcomes}}$, we first need to determine the total number of possible outfits. Once again, the Multiplication Rule applies; since there are 9 shirts, 7 pairs of pants, and 2 pairs of shoes, we can say there are $9 \cdot 7 \cdot 2 = 126$ total possible outfits. As we are looking for a blue shirt with black pants and brown shoes, the Multiplication Rule says there are $4 \cdot 2 \cdot 1 = 8$ ways this outfit could be selected. Therefore, the probability of Mr. Scott randomly choosing this outfit is $\frac{8}{126} = \frac{4}{63}$, or about **6.35%**.

Is the product of the individual probabilities equal to the total probability? The probability of selecting a white shirt is $\frac{4}{9}$, the probability of selecting black pants is $\frac{2}{7}$, and the probability of selecting brown shoes is $\frac{1}{2}$. The product of these probabilities is $\frac{4}{9} \cdot \frac{2}{7} \cdot \frac{1}{2} = \frac{8}{126}$, as anticipated. It seems we are ready to generalize.

PROBABILITY OF INDEPENDENT EVENTS

The probability of two (or more) independent events occurring together is the product of the probabilities of each individual event. In symbols, if there are two events A and B, this is written as $p(A \textit{ and } B) = p(A) \cdot p(B)$. If there are n events, this is written as $p(E_1 \textit{ and } E_2 \textit{ and } \ldots \textit{ and } E_n) = p(E_1) \cdot p(E_2) \ldots p(E_n)$.

The proof of this relies on the Multiplication Principle and fraction multiplication. We will prove this for two events, and the proof for n events is very similar and left to the reader.

PROOF

Let A and B be two independent events with probabilities $p(A)$ and $p(B)$, respectively. This means for each event there are some number of ways the event can occur and some total number of outcomes. In symbols, $p(A) = \frac{A_o}{A_t}$, where A_o represents the number of ways event A can occur out of a total number of outcomes A_t. Similarly, $p(B) = \frac{B_o}{B_t}$ where B_o represents the number of ways event B can occur out of a total number of outcomes B_t.

We now use the Multiplication Principle to determine how many total outcomes there are and how many different ways events A and B can occur. Since A and B are independent, we do not need to alter any of the values of A_o, A_t, B_o, or B_t. The Multiplication Principle tells us there are $A_t \cdot B_t$ total outcomes, and $A_o \cdot B_o$ different ways that events A and B can occur. Therefore, $p(A \text{ and } B) = \frac{A_o \cdot B_o}{A_t \cdot B_t}$. But $\frac{A_o \cdot B_o}{A_t \cdot B_t} = \frac{A_o}{A_t} \cdot \frac{B_o}{B_t}$, which equals $p(A) \cdot p(B)$, as desired.

Now that we have established the rule for the probability of independent events, we are able to calculate probabilities in situations where we may not be able to use $p(E) = \frac{\text{number of ways event } E \text{ can occur}}{\text{total number of outcomes}}$ because counting the number of ways an event can occur or the total number of outcomes is either impractical or impossible.

Let's consider an example. A manufacturer of laptop computers reports that only 5% of its laptops require significant maintenance within a year of purchase. If two laptops are purchased from this manufacturer, what is the probability that both will need significant maintenance within a year of purchase? Assuming the failure of each laptop to be an independent event, the probability of two laptops failing is $(0.05) \cdot (0.05) = 0.0025$, or 0.25%. This is an extremely unlikely event.

The assumption of independence in this situation is plausible, but by no means guaranteed. Maybe the defective laptops are disproportionately produced at a particular factory, or use a part from a different supplier from other laptops made by the company, so that the defective laptops are not evenly distributed throughout purchases. In this case, two laptops purchased together may be related (from a manufacturing standpoint), and so the failure of one may signal the increased likelihood of the other failing as well. The assumption of independence should always be questioned, as there may be times when assuming events are independent is not appropriate. For example, just because the probability of being struck by lightning is some value, say 0.002, doesn't mean the probability of an individual being struck by lightning twice is $(0.002) \cdot (0.002) = 0.000004$, because the events of the same person being struck by lightning twice may not be independent. Perhaps the individual works outdoors more frequently than most people, or maybe there is some particular physiological makeup that makes the individual more likely to be hit by lightning than others.

When the assumption of independence breaks down, we are not able to multiply the probabilities of the individual events together straight away. Next, we will look at some situations where the events are dependent and determine how we should approach these problems mathematically.

3.3.2 DEPENDENT EVENTS

Many events we experience appear to be independent, but often the probability of future events is dependent on the events that occur today. Although the identification of dependent events is usually clear upon reflection, sometimes, without careful consideration, an underlying assumption of independence may be missed. In this portion of the *Mathematics Resource Guide*, we will illustrate examples of when indepen dence should not be used and how to handle dependent problems mathematically. We will begin with an example which at first glance seems straightforward, but upon reflection is slightly more complicated than originally thought.

EXAMPLE 3.3L: The probability of winning the jackpot in a weekly lottery is $\frac{1}{60{,}000}$. If a person is selected at random from a phonebook and called, what is the probability that he/she is last week's winner?

SOLUTION: Although we want to say the probability of randomly selecting a winner is $\frac{1}{60{,}000}$ this is not the case. There are a few problems with this assumption. Some of the problems are logistical: what if the winner doesn't answer the phone or isn't listed in the phone book? What if the winner doesn't notice he/she has the winning lottery ticket and so never collects his/her prize? These are all valid concerns that interfere with the probability of selecting a winner being $\frac{1}{60{,}000}$.

However, even if all these concerns are dismissed, the probability of selecting a winner still isn't $\frac{1}{60{,}000}$. Let's say we are operating in a perfect logistical world: everyone is listed in the perfect phone book, and everyone always answers their phone, and every winner always collects his/her prize. Is the probability of selecting a winner $\frac{1}{60{,}000}$?

The probability of selecting a lottery winner at random from a phone book (even in a perfect logistical world) should be less than $\frac{1}{60{,}000}$ since not everyone plays the lottery. Say only 1 out of 3 people play the lottery. Then, in order to call someone at random and have them be a lottery winner, the person who is selected has to play the lottery, which isn't guaranteed to begin with. To select a person who played the lottery will take 3 calls (on average), and then only $\frac{1}{60{,}000}$ of those lottery players will be a winner. Therefore, the probability of randomly selecting a lottery winner from the phone book is $\frac{1}{3} \cdot \frac{1}{60{,}000} = \mathbf{\frac{1}{180{,}000}}$.

Mathematically this looks like what we do with independent events: the two probabilities are multiplied together. But, there is an important and subtle difference in what the $\frac{1}{60{,}000}$ represents. The $\frac{1}{60{,}000}$ is not the probability of winning the lottery; it is the probability of winning the lottery given that one purchases a ticket. The lottery does not select winners from people who don't purchase tickets, and one can imagine a lottery advertising campaign that tries to exploit this idea. ("I played the lottery, and I'm a winner. In fact, everyone who is a winner is someone who played the lottery.")

The events of purchasing a lottery ticket and winning the lottery are dependent events because the occurrence of one event (buying the ticket) affects the occurrence of the other event (winning the lottery). The probability of two dependent events occurring is the product of the probability the first event occurs and the probability the second event occurs given that the first event occurs.

PROBABILITY FOR DEPENDENT EVENTS

The probability of two dependent events occurring together is the product of the probability of the first event occurring and the probability of the second event occurring given the first event occurred. In symbols, if the two events are A and B, this is written as $p(A \text{ and } B) = p(A) \cdot p(B|A)$. $p(A)$ represents the probability of event A occurring, and $p(B|A)$, read aloud as "the probability of B given A," represents the probability that event B occurs if event A has already occurred. $p(B|A)$ is called a **conditional probability**.

The probability of winning the lottery is the probability of purchasing a ticket times the probability of winning given a ticket has been purchased. Let $p(T)$ represent the probability of purchasing a ticket and $p(W|T)$ be the probability of winning given a ticket has been purchased. Since $p(T) = \frac{1}{3}$ and $p(W|T) = \frac{1}{60{,}000}$, $p(T \text{ and } W) = p(T) \cdot p(W|T) = \frac{1}{3} \cdot \frac{1}{60{,}000} = \frac{1}{180{,}000}$.

Example 3.3m: What is the probability of purchasing a lottery ticket and not winning?

Solution: $p(T) = \frac{1}{3}$, and the probability of not winning given a ticket has been purchased is $p(W'|T) = \frac{59{,}999}{60{,}000}$. Therefore, $p(T \text{ and } W') = \frac{1}{3} \cdot \frac{59{,}999}{60{,}000} = \mathbf{\frac{59{,}999}{180{,}000}}$.

EXAMPLE 3.3N: What is the probability of not purchasing a lottery ticket and winning?

SOLUTION: Since $p(T) = \frac{1}{3}$, $p(T') = \frac{2}{3}$. However, given a ticket has not been purchased, it is impossible to win, so $p(W|T') = 0$. Therefore, $p(T' \text{ and } W) = \frac{2}{3} \cdot 0 = \mathbf{0}$.

EXAMPLE 3.3O: What is the probability of not purchasing a ticket and not winning?

SOLUTION: $p(T') = \frac{2}{3}$. Since not winning the lottery is a guarantee if no ticket is purchased, $p(W'|T') = 1$. Therefore $p(T' \text{ and } W') = \frac{2}{3} \cdot 1 = \mathbf{\frac{2}{3}}$.

Since these four scenarios (purchasing and winning, purchasing and not winning, not purchasing and winning, not purchasing and not winning) are the only four possible outcomes in this situation, the probabilities of these four events should sum to 1. Indeed, $\frac{1}{180{,}000} + \frac{59{,}999}{180{,}000} + 0 + \frac{2}{3} = 1$. Later on, we will discuss how we might use this distribution of probabilities to determine the best course of action with respect to this lottery.

EXAMPLE 3.3P: The probability of drawing a jack from a standard deck of playing cards is $\frac{4}{52} = \frac{1}{13}$. Is the probability of drawing two jacks from a deck of playing cards $\frac{1}{13} \cdot \frac{1}{13} = \frac{1}{169}$?

SOLUTION: It depends on whether the two events (drawing the first jack, drawing the second jack) are considered independent or dependent. If the first jack is replaced before the second jack is drawn, then the two events are independent, and the probability of drawing two jacks is $\mathbf{\frac{1}{169}}$. If, on the other hand, the two cards are drawn without replacement, then the probability is not $\frac{1}{169}$ because when the second jack is drawn, the probability of drawing a jack is no longer $\frac{1}{13}$. Once a jack is drawn, the probability of drawing another jack changes because now there are only 3 jacks out of 51 possible cards. Symbolically, we say $p(J|J) = \frac{3}{51}$, and therefore $p(J \text{ and } J) = p(J) \cdot p(J|J) = \frac{1}{13} \cdot \frac{1}{17} = \mathbf{\frac{1}{221}}$.

Example 3.3q: What is the probability of being dealt exactly three cards of the same value in a hand of five cards from a standard deck of playing cards?

Solution: Since this is a hand of cards, the cards are dealt without replacement, so the probabilities of later cards will depend on the cards previously dealt. Rather than picturing all five cards dealt simultaneously, let's imagine the cards being dealt one at a time. The first card can be any card, but the second and third card need to match that card, which means the probability of the second and third card being what we want are $\frac{3}{51}$ and $\frac{2}{50}$, respectively. Then, we need two additional cards that do not match the three we already have, so those probabilities are $\frac{48}{49}$ and $\frac{47}{48}$. Thus far we have $\frac{52}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{48}{49} \cdot \frac{47}{48} = \frac{47}{20825}$.

But, this calculation assumes we were dealt the five cards in a specific order: *matching, matching, matching, something else, something else*. We don't necessarily need the cards to be in this order; we just need these cards to be in the hand at the end of the deal. Therefore, we need to consider the different possible arrangements of the five cards. If we start to list all the possibilities, this looks like:

match, match, match, else, else

match, match, else, match, else

match, else, match, match, else

But wait! We have five spots, and we need to select three for the matching cards. No spot can be selected more than once, and the order of selection does not matter. This is a combination! Therefore, there are $\binom{5}{3} = 10$ ways this can occur.

Therefore, the final probability of being dealt exactly three cards of the same value in a hand of five cards is $\frac{52}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{48}{49} \cdot \frac{47}{48} \cdot \binom{5}{3} = \frac{94}{4165}$, or approximately **2.257%**.

We will conclude this section with one more example using probability of dependent events.

The probability of getting a 34 or higher on the ACT math section is 8%, and the probability of getting 700 or higher on the SAT math section is 6%. But for students who took both tests, the probability of getting a 34 or higher on the ACT math section and 700 or higher on the SAT math section is 5.5%. How is this possible? If you knew your friend had an ACT math score of 35, what is the probability that your friend also scored higher than 700 on the SAT math section?

First we introduce a little notation. Let $p(ACT)$ and $p(SAT)$ represent the probabilities of scoring 34 or higher and 700 or higher on the math sections of the ACT and SAT, respectively. Therefore $p(ACT) = 0.08$ and $p(SAT) = 0.06$.

Yet $(0.08) \cdot (0.06) = 0.0048$, which implies that 0.48% of students who take both tests should score above the thresholds on both tests. This is vastly different from the actual probability of 5.5%. What accounts for this differential?

$p(ACT\ and\ SAT) = p(ACT) \cdot p(SAT)$ only if the two events are independent, but we anticipate there should be a great deal of crossover between students who scored well on the math section of the ACT and students who scored well on the math section of the SAT. Therefore, these events are dependent, and $p(ACT\ and\ SAT) = p(ACT) \cdot p(SAT|\mathrm{ACT})$. Substituting known values yields $0.055 = (0.08) \cdot p(SAT|ACT)$, and so $p(SAT|ACT) = 0.6875$. This means the probability of scoring 700 or higher on the SAT math section given a score of 34 or higher on the ACT math section is 68.75%. Therefore, if your friend scored 34 or higher on the ACT math section but below 700 on the SAT math section, we might be somewhat surprised, but not totally shocked.

Notice that $p(SAT|ACT) \neq p(ACT|SAT)$. $p(ACT\ and\ SAT) = p(SAT) \cdot p(ACT|SAT)$ and so $0.055 = (0.06) \cdot p(ACT|SAT)$ and $p(ACT|SAT) = 0.9167$. This means the probability of scoring 34 or higher on the ACT math section given a score of 700 or higher on the SAT math section is 91.67%, so a high score on the SAT math section without a high score on the ACT math section is fairly rare.

The differences between $p(ACT\ and\ SAT)$, $p(ACT)$, $p(SAT)$, p($SAT|ACT$), and $p(ACT|SAT)$ may be more clear if we consider a table showing the breakdown of possible scenarios. Assume 100 people took both the ACT and SAT. Based on our given information, 8 people will have scored 34 or higher on the ACT math section, 6 will have scored 700 or higher on the SAT math Section, and 5.5 people will have scored higher than both of these cutoffs.

	≥700 SAT Math	< 700 SAT Math	**TOTAL**
≥ 34 ACT Math	5.5		**8**
< 34 on ACT Math			
TOTAL	**6**		**100**

Fortunately, we have enough information to complete the entire table.

	≥700 SAT Math	< 700 SAT Math	TOTAL
≥ 34 ACT Math	5.5	2.5	**8**
< 34 on ACT Math	0.5	91.5	**92**
TOTAL	**6**	**94**	**100**

Now the difference between $p(SAT|ACT)$ and $p(ACT|SAT)$ may be more clear. When we calculate $p(SAT|ACT)$, the probability of a high score on the SAT given a high score on the ACT, we are considering only students who already had a high score on the ACT math section, so the probability is $\frac{5.5}{8} = 0.6875$. When we calculate $p(ACT|SAT)$, the probability of a high score on the ACT given a high score on the SAT, we are considering only students who already had a high score on the SAT math section, so the probability is $\frac{5.5}{6} = 0.9167$.

3.4 PROBABILITY DISTRIBUTIONS

Once we are comfortable calculating the probability of a specific outcome within a scenario, we turn our attention to considering all of the possible outcomes for a given situation. When making decisions, we are usually confronted with an array of possible outcomes, each of which has a particular value and a particular probability. How can we determine which course of action is the most mathematically sound? In this section, we will consider how to determine the appropriate choice given events with different probabilities and different outcomes.

In order to be a probability distribution, all possible outcomes must be represented with a valid probability. This means the distribution must satisfy two of the basic probability principles:

(1) The probability of each event is a number between 0 and 1.

(2) The sum of all probabilities must equal 1.

DEFINITION OF A PROBABILITY DISTRIBUTION

A probability distribution is a set of outcomes and the probability of those outcomes where the probability of each event is a number between 0 and 1, and the sum of the probabilities from all outcomes equals 1. Symbolically, if there are m different outcomes, for all events E_i within the probability distribution, $0 \le p(E_i) \le 1$ and $\sum_{i=1}^{m} p(E_i) = 1$.

As an example, we can build a probability distribution for the lottery situation from the previous section. It was determined that the probability of buying a ticket and winning the lottery was $\frac{1}{180,000}$, the probability of buying a ticket and not winning the lottery was $\frac{59,999}{180,000}$, the probability of not buying a ticket and winning the lottery was 0, and the probability of not buying a ticket and not winning the lottery was $\frac{2}{3}$. As each probability is between 0 and 1 (inclusive) and the sum of all the probabilities is 1, this is a valid probability distribution. Often we display probability distributions in tables.

E_i	*ticket and win*	*ticket and not win*	*no ticket and win*	*no ticket and not win*
$P(E_i)$	$\frac{1}{180,000}$	$\frac{59,999}{180,000}$	0	$\frac{2}{3}$

But, knowing the probabilities does not give us enough information to make a mathematically sound decision. We also need to know the outcome associated with each event; it should make a difference if a winning lottery ticket is worth $10 or $1,000. This leads us to our discussion of **expected value.**

3.4.1 EXPECTED VALUE

Suppose the lottery ticket costs $3, but a winning lottery ticket is worth $15,000. Is it worth playing this lottery? Paying $3 for a chance at $15,000 seems like a fairly tempting lottery ticket, but the probabilities of winning and losing need to be taken into account as well.

Let's pretend we simulate this scenario many, many times—say 360,000 times. What do we anticipate occurring? If we decided whether or not to play the lottery 360,000 times, we would expect to have 2 winning lottery tickets to go along with 119,998 losing lottery tickets. Additionally, we would have decided not to play the lottery 240,000 times.

E_i	*ticket and win*	*ticket and not win*	*no ticket and win*	*no ticket and not win*
$P(E_i)$	$\frac{1}{180,000}$	$\frac{59,999}{180,000}$	0	$\frac{2}{3}$
# of times	2	119,998	0	240,000
PAYOFF	**$29,994**	**$–359,994**	**$0**	**$0**

Two winning lottery tickets worth $30,000 (less the $6 these two tickets cost) seems really nice, until we realize that we lost over $350,000 in order to purchase those two lottery tickets. It turns out this is a terrible lottery to play.

Using this information, we can determine the value of an average lottery decision. Since all 360,000 decisions had a total value of \$–330,000, the mean value for a decision is $\frac{\$-330{,}000}{360{,}000} = \$-0.91667$. Therefore, on average each time a decision is made in this scenario, the outcome is to lose about 92 cents. This average outcome is called the **expected value**.

DEFINITION OF EXPECTED VALUE

The **expected value** of a probability distribution is the mean payoff expected from the probability distribution over a long enough period of time.

Like the mean of a data set, it is possible for the expected value to not be a possible outcome. In the lottery example, there are only three possible outcomes for an individual decision: \$15,000, \$–3, or \$0. The expected value of \$–0.916 is also based on the decision of whether or not to play the lottery being random, and that a ticket is purchased $\frac{1}{3}$ of the time and not purchased $\frac{2}{3}$ of the time. Most people do not randomly decide whether to play the lottery. In this situation, since one can control whether or not to purchase a ticket, the choice made should always be to not buy a ticket, as this creates an outcome higher than the expected value.

For this lottery, how much does a winning lottery ticket need to be worth for it to be worthwhile for people to choose to play the lottery? Randomly deciding whether or not to play the lottery 360,000 times resulted in a total cost of \$359,994 from losing lottery tickets and only two winning tickets. In order to make people want to play the lottery, the two winning tickets need to be worth more than \$359,994 combined. So, if a winning lottery ticket gains the purchaser \$179,997, the expected value for this situation is \$0, and it does not matter if we purchase a lottery ticket or not. Therefore, a winning lottery ticket needs to be worth more than \$180,000 (remember they still cost \$3) in order to create a positive expected value. At that point, people should choose to play the lottery because they expect to get back more money than they spend in the long run.

This example illustrates the concept of a **fair game**.

DEFINITION OF A FAIR GAME

A **fair game** is a probability distribution with an expected value of 0.

Keep in mind that the expected value of a probability distribution is the average outcome over a long period of time with many, many trials. Expected values that are slightly negative still represent very large losses over a long period of time. Lotteries and casinos will set up games with a slightly negative expected value. Although this means some players may make money in the short run, in the long run the casinos will take in more money than they pay out. Indeed, allowing a few players to win money probably will cause more people to gamble more often!

So, how do we calculate expected value? Let's look at another example with an eye toward generalizing the process to find a formula. A local casino introduces a new game, *Snake Eyes*, which is advertised for its ease of play. The gambler pays \$2 to roll a pair of dice numbered 1–6. The payout is based on the sum of the two dice as follows:

A sum of 2 pays the gambler \$100.

A sum of 10 gives the gambler \$2 (so the play was free).

A sum of 11 gives the gambler \$10.

A sum of 7 causes the gambler to have to pay \$1 *more*.

All other sums result in nothing happening (the gambler effectively loses \$2).

Should gamblers play this game?

To determine whether or not this game should be played, we need to find the expected value. A positive expected value means that over time, gamblers will win more money than they lose, whereas a negative expected value means gamblers will lose more money than they win. Let's imagine the game being played a large number of times, and let's select a multiple of 36, so our calculations come out nicely. If this game is played 720 times, what should occur?

A sum of 2 should occur $\frac{1}{36} \cdot 720 = 20$ times, for a payout of $20 \cdot 98 = \$1{,}960$. (Don't forget that it costs \$2 to play!)

A sum of 10 should occur $\frac{3}{36} \cdot 720 = 60$ times, for a payout of \$0.

A sum of 11 should occur $\frac{2}{36} \cdot 720 = 40$ times, for a payout of $40 \cdot 8 = \$320$.

A sum of 7 should occur $\frac{6}{36} \cdot 720 = 120$ times, for a payout of $-\$360$.

All other sums should occur $\frac{24}{36} \cdot 720 = 480$ times, for a payout of $480 \cdot (-2) = -\$960$.

All together, the expected payout is \$1,960 + \$0 + \$320 + (–\$360) + (–\$960) = \$960. As it took 720 turns to accumulate this amount, the expected value is $\frac{960}{720}$ = \$1.33 per play. This is an extremely good expected value for a gambler, and someone at the casino is probably going to lose his/her job for designing a game with such a high expected value.

How did we find this expected value? We first determined the expected occurrence of each outcome by multiplying the probability of that outcome by the total number of plays, in this case 720. Then, each number of occurrences was multiplied by the payout for that outcome, and all of these total payouts were added together. Then, we divided the total amount of money by 720, the total number of plays, to determine the average payout for one game. These calculations, without evaluating as we go, look like this:

$$\frac{\frac{1}{36}\cdot 720\cdot 98+\frac{3}{36}\cdot 720\cdot 0+\frac{2}{36}\cdot 720\cdot 8+\frac{6}{36}\cdot 720\cdot(-3)+\frac{24}{36}\cdot 720\cdot(-2)}{720}.$$

But, each of the terms in the numerator contains a factor of 720, which can be factored out to give us:

$$\frac{720\cdot\left[\frac{1}{36}\cdot 98+\frac{3}{36}\cdot 0+\frac{2}{36}\cdot 8+\frac{6}{36}\cdot(-3)+\frac{24}{36}\cdot(-2)\right]}{720}=\frac{1}{36}\cdot 98+\frac{3}{36}\cdot 0+\frac{2}{36}\cdot 8+\frac{6}{36}\cdot(-3)+\frac{24}{36}\cdot(-2).$$

But, this is just the probability of each outcome multiplied by the value of the outcome, and then the sum of these products. Indeed, for any number of trials, we would multiply the number of trials by the probability of each outcome to find the number of occurrences, and then multiply the number of occurrences by the value of these outcomes. This would be done for each outcome, and then we would add together the results and divide by the number of outcomes to find the average. Symbolically, if n represents the number of trials and there are m different possible outcomes, we have:

$$\frac{p(E_1)\cdot n\cdot E_1+p(E_2)\cdot n\cdot E_2+p(E_3)\cdot n\cdot E_3+\ldots+p(E_m)\cdot n\cdot E_m}{n}.$$

Since there is a factor of n in all of these terms, these can be canceled, and the computation can be written as: $p(E_1)\cdot E_1+p(E_2)\cdot E_2+p(E_3)\cdot E_3+\ldots+p(E_m)\cdot E_m$.

So, the expected value of a probability distribution is a weighted sum of all the possible outcomes, where the weight of each outcome is the probability of that outcome. But, this also looks a lot like a series! We can use sigma notation to write the final version of the formula.

FORMULA FOR THE EXPECTED VALUE OF A PROBABILITY DISTRIBUTION

The expected value of a probability distribution with m different outcomes $E_1, E_2, \ldots E_m$, each with a corresponding probability of occurrence $p(E_1), p(E_2), \ldots p(E_m)$, has expected value $\sum_{i=1}^{m} p(E_i) \cdot E_i$. The symbol we will use to denote the expected value of a probability distribution is $\overline{E}$.

Let's use this formula to determine the expected value of the lottery situation from the beginning of our discussion. We already know the expected value of the probability distribution is \$–0.91667, so our new formula needs to give us the same value in order for us to feel confident that this formula is correct.

In our lottery situation, there are four possible outcomes. The first outcome, E_1, is buying a ticket and winning the lottery, so $E_1 = 14{,}997$. This has a probability of $\frac{1}{180{,}000}$.

The second outcome, E_2, is buying a ticket and not winning the lottery, so $E_2 = -3$. This has a probability of $\frac{59{,}999}{180{,}000}$.

The third event, not buying a ticket and winning the lottery, has a probability of 0. Since this event cannot occur, we will omit this from our probability distribution.

The third outcome, E_3, is not buying a ticket and not winning the lottery, so $E_3 = 0$. This has a probability of $\frac{2}{3}$.

Using our formula for expected value, $\overline{E} = \sum_{i=1}^{m} p(E_i) \cdot E_i$, gives us $\frac{1}{180{,}000} \cdot 14997 + \frac{59{,}999}{180{,}000} \cdot -3 + \frac{2}{3} \cdot 0 = -0.91667$, as predicted! So, our expected value formula works the way we want it to, and measures what we set out to measure.

Having a measure of central tendency for a probability distribution is very nice, but as soon as we can measure the mean of a probability distribution, we start to wonder: can we measure the variation of a probability distribution as well?

3.4.2 VARIANCE AND STANDARD DEVIATION OF PROBABILITY DISTRIBUTIONS

Expected value tells us the average value of a probability distribution and can help us determine whether or not to participate in a particular probability situation, like a lottery. If the expected value is positive, then in the long run we expect to have a positive return. Lotteries and casinos make sure the expected values of their games are negative, so they take in more money than they pay out. But probability distributions contain variation, just like collections of data. Being able to quantify this variation can help clever gamblers select which games to play since games with negative expected value but larger variations could create positive payoffs if played for a short period of time.

It seems the variance of a probability distribution should be calculated in a similar manner to the variance of a data set. Variance measures the average of the squared differences each value is from the mean: $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}$. The variance of a probability distribution would need to take the probability of each outcome into account as well. So, a probability distribution with m outcomes $E_1, E_2, \ldots E_m$, each probability $p(E_1), p(E_2), \ldots p(E_m)$, and expected value $\overline{E}$ should have variance $\sigma^2 = (E_1 - \overline{E})^2 \cdot p(E_1) + (E_2 - \overline{E})^2 \cdot p(E_2) + \ldots + (E_m - \overline{E})^2 \cdot p(E_m)$. This can be written in sigma notation as $\sigma^2 = \sum_{i=1}^{m}(E_i - \overline{E})^2 \cdot p(E_i)$.

VARIANCE OF A PROBABILITY DISTRIBUTION

The variance of a probability distribution is the sum of the squares of the differences between each outcome and the expected value multiplied by the probability of each outcome. Symbolically, $\sigma^2 = \sum_{i=1}^{m}(E_i - \overline{E})^2 \cdot p(E_i)$.

As we might expect from our earlier work with variance and standard deviation, the standard deviation of a probability distribution is the square root of the variance.

STANDARD DEVIATION OF A PROBABILITY DISTRIBUTION

The standard deviation of a probability distribution is the square root of the variance of a probability distribution. Symbolically, $\sigma = \sqrt{\sum_{i=1}^{m}(E_i - \overline{E})^2 \cdot p(E_i)}$.

EXAMPLE 3.4A: The expected value of the lottery situation from the previous section was \$–0.91667. Find the variance and standard deviation for this probability distribution.

SOLUTION: For the lottery situation, $E_1 = 14{,}997$, so $E_1 - \overline{E} = 14{,}997.91667$, and $(E_1 - \overline{E})^2 = 224{,}937{,}504.44$. This has a probability of $\frac{1}{180{,}000}$ and so $(E_1 - \overline{E})^2 \cdot p(E_1) = 1249.653$.

$E_2 = -3$, so $E_2 - \overline{E} = -2.0833$, and $(E_2 - \overline{E})^2 = 4.3403$. This has a probability of $\frac{59{,}999}{180{,}000}$ and so $(E_2 - \overline{E})^2 \cdot p(E_2) = 1.4467$.

$E_3 = 0$, so $E_3 - \overline{E} = 0.91667$, and $(E_3 - \overline{E})^2 = 0.8402$. This has a probability of $\frac{2}{3}$, and so $(E_3 - \overline{E})^2 \cdot p(E_3) = 0.5601$.

Therefore, the variance for this probability distribution is $1249.653 + 1.4467 + 0.5601 = 1251.6598$, and the standard deviation is $\sqrt{1251.6598} = \mathbf{35.378}$.

Why would we want to calculate the variance and standard deviation for a probability distribution? One reason is that once we know the mean, variance, and standard deviation for two probability distributions, we can compare the relative likelihood of two outcomes in different probability distributions using z-scores in the same way we used z-scores to compare the relative position of two data points in two different sets of data.

EXAMPLE 3.4B: What is the z-score of winning the lottery in Example 3.4a? What is the z-score of playing the lottery and not winning?

SOLUTION: With a standard deviation of 35.378 and an expected value of –0.91667, an outcome of winning \$14,997 has the incredibly high z-score of $\frac{14{,}997 - (-0.91667)}{35.378} = 423.933$. Not winning the lottery, on the other hand, has a much more reasonable z-score of $\frac{-3 - (-0.91667)}{35.378} = \mathbf{-0.059}$.

EXAMPLE 3.4C: What is the variance and standard deviation for Snake Eyes as described in the previous section?

SOLUTION: Recall the rules for Snake Eyes: the gambler pays \$2 and rolls a pair of dice. The sum of the two dice determines what happens:

A sum of 2 pays the gambler \$100. This has a probability of $\frac{1}{36}$.

A sum of 10 gives the gambler \$2 (so the play was free). This has a probability of $\frac{3}{36}$.

A sum of 11 gives the gambler \$10. This has a probability of $\frac{2}{36}$.

A sum of 7 causes the gambler to have to pay \$1 *more*. This has a probability of $\frac{6}{36}$.

All other sums result in nothing occurring (so the gambler loses \$2). This has a probability of $\frac{24}{36}$.

The expected value for this game was previously determined to be 1.33. Therefore, the variance for this probability distribution is:

$$(98 - 1.33)^2 \cdot \frac{1}{36} + (0 - 1.33)^2 \cdot \frac{3}{36} + (8 - 1.33)^2 \cdot \frac{2}{36} + [(-3) - 1.33]^2 \cdot \frac{6}{36} + [(-2) - 1.33]^2 \cdot \frac{24}{36} = \mathbf{272.722}.$$

The standard deviation for this probability distribution is therefore $\sqrt{272.722} = \mathbf{16.514}$.

The z-score for playing *Snake Eyes* and winning \$98 is therefore $\frac{98 - 1.33}{16.514} = 5.854$, still a fairly high z-score, but far better than the astronomical z-score for winning the lottery of 423.933. The calculations for standard deviation and expected value take the probability of winning and the payout into account as well, so even with these factors controlled for, playing *Snake Eyes* and hoping for the best payout is a far more reasonable thing to do than playing the lottery and hoping for the \$15,000 payoff.

Unlike mean, variance, and standard deviation for data, most calculators and computer programs do not have a command or shortcut for calculating the expected value, variance, and standard deviations of probability distributions. Although calculating expected value using a calculator or spreadsheet is fairly straightforward, the variance formula we just used is somewhat cumbersome. Like the variance formula for data sets, there is an alternative form for the variance formula that may be somewhat easier to use for calculations. We will conclude this section with its derivation.

Consider $\sigma^2 = \sum_{i=1}^{m} (E_i - \overline{E})^2 \cdot p(E_i)$, where $\overline{E}$ is the expected value of the probability distribution. Writing out a few terms to get a feel for this calculation gives us: $(E_1 - \overline{E})^2 \cdot p(E_1) + (E_2 - \overline{E})^2 \cdot p(E_2) + (E_3 - \overline{E})^2 \cdot p(E_3) + \ldots + (E_m - \overline{E})^2 \cdot p(E_m)$.

Expanding each of the binomials and distributing each $p(E_i)$ gives us what looks like a mess at first:

$$E_1{}^2 \cdot p(E_1) - 2E_1\overline{E} \cdot p(E_1) + \overline{E}^2 \cdot p(E_1) + E_2{}^2 \cdot p(E_2) - 2E_2\overline{E} \cdot p(E_2) + \overline{E}^2 \cdot p(E_2)$$

$$+ E_3{}^2 \cdot p(E_3) - 2E_3\overline{E} \cdot p(E_3) + \overline{E}^2 \cdot p(E_3) + \ldots + E_m{}^2 \cdot p(E_m) - 2E_m\overline{E} \cdot p(E_m) + \overline{E}^2 \cdot p(E_m)$$

This looks pretty ugly. Let's regroup and place all the terms that look like $E_1{}^2 \cdot p(E_1)$ together, all the terms that look like $-2E_1\overline{E} \cdot p(E_1)$together, and all the terms that look like $\overline{E}^2 \cdot p(E_1)$ together:

$$E_1{}^2 \cdot p(E_1) + E_2{}^2 \cdot p(E_2) + E_3{}^2 \cdot p(E_3) + \ldots + E_m{}^2 \cdot p(E_m)$$

$$- 2E_1\overline{E} \cdot p(E_1) - 2E_2\overline{E} \cdot p(E_2) - 2E_3\overline{E} \cdot p(E_3) - \ldots - 2E_m\overline{E} \cdot p(E_m)$$

$$+ \overline{E}^2 \cdot p(E_1) + \overline{E}^2 \cdot p(E_2) + \overline{E}^2 \cdot p(E_3) + \ldots + \overline{E}^2 \cdot p(E_m)$$

Now let's try to simplify using sigma notation. The first set can be written as $\sum_{i=1}^{m} E_i{}^2 \cdot p(E_i)$. In the second set, each term contains a common factor of –2 and $\overline{E}$, so once these are factored out, this can be written as $-2\overline{E} \cdot \sum_{i=1}^{m} E_i \cdot p(E_i)$. In the third set, each term contains a common factor of $\overline{E}^2$, and so these terms can be written as $\overline{E}^2 \cdot \sum_{i=1}^{m} p(E_i)$.

Therefore, we have: $\sigma^2 = \sum_{i=1}^{m} (E_i - \overline{E})^2 \cdot p(E_i)$

$$= \sum_{i=1}^{m} E_i^2 \cdot p(E_i) - 2\overline{E} \cdot \sum_{i=1}^{m} E_i \cdot p(E_i) + \overline{E}^2 \cdot \sum_{i=1}^{m} p(E_i).$$

But $\sum_{i=1}^{m} E_i \cdot p(E_i)$ is the expected value of the probability distribution, so $\sum_{i=1}^{m} E_i \cdot p(E_i) = \overline{E}$. And $\sum_{i=1}^{m} p(E_i)$ is asking for the sum of all the probabilities in the distribution, which we know equals 1! Therefore, this simplifies further, to:

$$= \sum_{i=1}^{m} E_i^2 \cdot p(E_i) - 2\overline{E}^2 + \overline{E}^2 = \sum_{i=1}^{m} [E_i^2 \cdot p(E_i)] - \overline{E}^2.$$

ALTERNATIVE FORMULA FOR THE VARIANCE OF A PROBABILITY DISTRIBUTION

The variance of a probability distribution with m outcomes $E_1, E_2, \ldots E_m$, each with a corresponding probability of occurrence $p(E_1), p(E_2), \ldots p(E_m)$ and an expected value of $\overline{E}$ can be calculated using the following formula: $\sigma^2 = \sum_{i=1}^{m} [E_i{}^2 \cdot p(E_i)] - \overline{E}^2$.

Probability distributions are an extremely important foundational concept in statistics. Many different types of probability distributions are studied and developed in statistics, but all probability distributions are governed by the rules and formulas we have discussed thus far. Now that we have some familiarity with probability distributions, we will turn our attention to two of the most important probability distributions in statistics: the Binomial Distribution and the Normal Distribution.

3.5 THE BINOMIAL DISTRIBUTION

The Binomial Distribution is an important probability distribution in statistics because it lays the groundwork for the Normal Distribution. It turns out, perhaps not surprisingly, that the Binomial Distribution in statistics is closely related to the algebraic Binomial Expansion Theorem. In this section of the *Mathematics Resource Guide*, we will develop the Binomial Distribution, consider its relationship to the Binomial Expansion Theorem, and determine the expected value, variance, and standard deviation of the Binomial Distribution. In the last section of the resource guide, we will consider how the Binomial Distribution leads into the Normal Distribution.

Let's consider an example. On your way into biology class, you suddenly remember there is a 10-question quiz today on the assigned reading. Having completely forgotten about the quiz, you are forced to guess on all 10 questions. Each question is multiple-choice with four possible choices, one of which is correct. What is the probability that you will pass the quiz with a 60% or better?

The probability of guessing any one of the questions correctly is 25%, or 0.25, and the probability of getting a question incorrect is 0.75. Since the probability of guessing the first question correctly is independent of guessing the second question correctly, we can multiply the probabilities we want together (without having to worry about conditional probabilities). So, the probability of getting exactly 6 of the 10 questions correct seems to be $(0.25) \cdot (0.25) \cdot (0.25) \cdot (0.25) \cdot (0.25) \cdot (0.25) \cdot (0.75) \cdot (0.75) \cdot (0.75) \cdot (0.75) = (0.25)^6 \cdot (0.75)^4$. But, this assumes that the first six questions were answered correctly and the last four questions were answered incorrectly, which does not have to be the case. Any six of the ten questions could have been answered correctly, and any four of the ten could be answered incorrectly. So, clearly there are more possibilities than *ccccccwwww*. A few others could be:

cccccwcwww

ccwwccwwcc

wcwcwcwccc

This lineup representation looks familiar! As long as six of the ten spots are selected for c (getting that question correct), then this is a valid set. The spots are selected without replacement, and the order of selection doesn't matter, so this is a combination!

Therefore, there are $\binom{10}{6} = 210$ different ways you could guess 6 of the 10 questions on the quiz correctly, and the total probability of getting 6 out of the 10 questions correct is $\binom{10}{6} \cdot (0.25)^6 \cdot (0.75)^4 = 0.01622$, or 1.622%. This doesn't seem like a very promising chance to pass the quiz.

But wait! This is the probability of getting exactly 6 of the 10 questions correct. Getting more than 6 questions correct also counts as a passing grade! So, you could also get 7, 8, 9, or 10 questions correct.

Fortunately, these probabilities are very close in structure to the probability of getting 6 out of 10 correct.

Unfortunately, none of these outcomes are very likely.

The probability of getting 7 out of 10 correct is $\binom{10}{7} \cdot (0.25)^7 \cdot (0.75)^3 = 0.00309$.

The probability of getting 8 out of 10 correct is $\binom{10}{8} \cdot (0.25)^8 \cdot (0.75)^2 = 0.00038$.

The probability of getting 9 out of 10 correct is $\binom{10}{9} \cdot (0.25)^9 \cdot (0.75)^1 = 0.000028$.

And, the probability of getting all 10 correct is the astronomically low $\binom{10}{10} \cdot (0.25)^{10} \cdot (0.75)^0 = .00000095$.

Therefore, the probability of getting 6 or more questions correct out of 10 is the sum of these probabilities, or 0.019727. So, there is only a 1.9727% chance of passing the quiz when guessing on every problem. Ouch! Better remember to study for the quiz next time!

Hopefully the similarity to the Binomial Expansion Theorem is clear, as each probability resembles a term in the expanded form of $(x + y)^n$: $\binom{n}{k} \cdot x^k \cdot y^{n-k}$. For this example, $n = 10$, $x = 0.25$, $y = 0.75$, and k is ranging from 6 to 10. Notice that $x + y = 1$ since each question is either guessed correctly or incorrectly. This is an important characteristic of the Binomial Distribution—each event can be classified as having two possible outcomes: success or failure. In this example, even though there are four choices for each multiple-choice question, each question is scored right or wrong.

Let's look at another example before we generalize.

EXAMPLE 3.5A: In your math class, students are arranged into groups of four, and each day homework points are awarded to all members of a group or no members of a group. In order for the group to receive credit for the homework assignment, at least three members of the group need to complete the daily assignment. Assuming each group member has a 70% chance of completing the assignment on any given day, what is the probability a group earns credit for the homework on a single day?

SOLUTION: In order for the group to earn credit on the homework, either three or four of the group members need to complete the assignment. The probability of three group members completing the assignment is $\binom{4}{3} \cdot (0.7)^3 \cdot (0.3)^1 = 0.4116$, and the probability of all four group members completing their assignment is $\binom{4}{4} \cdot (0.7)^4 \cdot (0.3)^0 = 0.2401$. Therefore, the probability of either three or four group members completing the daily assignment is 0.6517, and there is a **65.17%** chance that a group will earn credit for the homework on any given day.

Based on our examples, it seems we are ready to generalize.

CRITERIA AND FORMULA FOR THE BINOMIAL DISTRIBUTION

A situation can be modeled using the Binomial Distribution if the following criteria hold: each outcome is made up of a series of independent events, these events can be classified as a success or a failure, and the probability of success is constant across events. Given a probability of success p and a probability of failure q for each event, the probability of k successes out of n events is given by the formula $P(k \text{ successes out of } n \text{ events}) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$. Note that since p is the probability of success and q is the probability of failure, $p + q = 1$. Also notice that with n events, there are $n + 1$ different possible outcomes, since it is possible for the event to occur 0 out of n times.

The similarity to the Binomial Expansion Theorem is hopefully clear. Indeed, we can use this connection to the Binomial Expansion Theorem to prove that the Binomial Distribution is a probability distribution.

In order to be a probability distribution, the Binomial Distribution must satisfy two important properties: each outcome must have a probability between 0 and 1 (inclusive), and the sum of the probabilities of all outcomes must be 1. Symbolically, $0 \leq \binom{n}{k} \cdot p^k \cdot q^{n-k} \leq 1$ and $\sum_{i=0}^{n} \binom{n}{k} \cdot p^k \cdot q^{n-k} = 1$. We will prove the latter condition first, and then we will use that result to prove the first condition.

In order to show that $\sum_{i=0}^{n} \binom{n}{k} \cdot p^k \cdot q^{n-k} = 1$, we use the Binomial Expansion Theorem: $(x + y)^n = \sum_{i=0}^{n} \binom{n}{k} \cdot x^k \cdot y^{n-k}$. Substituting p for x and q for y gives us $(p + q)^n = \sum_{i=0}^{n} \binom{n}{k} \cdot p^k \cdot q^{n-k}$. But, since p represents the probability of success and q represents the probability of failure, $p + q = 1$. Since $1^n = 1$, we have $1 = \sum_{i=0}^{n} \binom{n}{k} \cdot p^k \cdot q^{n-k}$ as desired.

Now that we know that the sum of all probabilities from all outcomes from the Binomial Distribution is 1, we will use this to show that each individual probability is between 0 and 1. First we will show that all the probabilities are non-negative. Since each probability is given by $\binom{n}{k} \cdot p^k \cdot q^{n-k}$, we can determine that this probability is non-negative by showing that none of the three terms in the product are negative. $\binom{n}{k}$ relies on a computation from factorials, none of which will be negative, and p and q are probabilities, which will also be non-negative. Since an integer power of a non-negative number remains non-negative, both p^k and q^{n-k} are non-negative, and the product of three non-negative numbers is also non-negative.

This is true for all outcomes in a given Binomial Distribution. But, we know that the sum of all possible outcomes for a Binomial Distribution is 1, and if each term is non-negative, this implies that each of the terms is also less than 1, as desired. (If the sum of all the terms is 1, the only way one or more of them could be larger than 1 is to have some negative terms to lower the sum back to 1, but we have shown none of the terms are negative.) Therefore, the Binomial Distribution is a probability distribution.

THE BINOMIAL DISTRIBUTION IS A PROBABILITY DISTRIBUTION

It has been proven that the Binomial Distribution is a probability distribution.

If the Binomial Distribution is a probability distribution, we should be able to find the expected value, variance, and standard deviation for the Binomial Distribution.

EXAMPLE 3.5B: Find the expected value, variance, and standard deviation of the number of heads that will occur if 8 fair coins are flipped.

SOLUTION: First we note that this situation can be modeled using the Binomial Distribution since each outcome (getting a certain number of heads) is made up of independent events (each coin is independent of the others), the events can be modeled with success or failure (heads is success, tails is failure), and the probability of success remains constant across events (the probability of getting a head on each coin is $\frac{1}{2}$). In order to find the expected value, we need to determine each outcome and the probability of each outcome.

# OF HEADS	BINOMIAL PROBABILITY	DECIMAL PROBABILITY	$p(E_i) \cdot E_i$	$E_i^2 \cdot p(E_i)$
0	$\binom{8}{0} \cdot (.5)^0 \cdot (.5)^8$	0.00390625	0	0
1	$\binom{8}{1} \cdot (.5)^1 \cdot (.5)^7$	0.03125	0.03125	0.03125
2	$\binom{8}{2} \cdot (.5)^2 \cdot (.5)^6$	0.109375	0.21875	0.4375
3	$\binom{8}{3} \cdot (.5)^3 \cdot (.5)^5$	0.21875	0.65625	1.96875
4	$\binom{8}{4} \cdot (.5)^4 \cdot (.5)^4$	0.273438	1.09375	4.375
5	$\binom{8}{5} \cdot (.5)^5 \cdot (.5)^3$	0.21875	1.09375	5.46875
6	$\binom{8}{6} \cdot (.5)^6 \cdot (.5)^2$	0.109375	0.65625	3.9375
7	$\binom{8}{7} \cdot (.5)^7 \cdot (.5)^1$	0.03125	0.21875	1.53125
8	$\binom{8}{8} \cdot (.5)^8 \cdot (.5)^0$	0.00390625	0.03125	0.25
TOTAL		**1**	**4**	**18**

Not surprisingly, the expected value of this probability distribution is 4, since we would expect to get 4 heads when we flip 8 coins "on average."

This idea leads us to hypothesize the following result.

EXPECTED VALUE OF A BINOMIAL DISTRIBUTION

The expected value of a binomial distribution with n events and probability of success p on each event is $\overline{E} = n \cdot p$.

The last column in the table for Example 3.5B shows $E_i^2 \cdot p(E_i)$, which is used in the alternative form of the variance formula for a probability distribution. Since $\sum_{i=0}^{m} E_i^2 \cdot p(E_i) = 18$ and $\overline{E} = 4$, the variance of this probability distribution is $18 - 4^2 = 2$, and therefore the standard deviation of this probability distribution is $\sqrt{2}$. This leads us to hypothesize the following result.

VARIANCE AND STANDARD DEVIATION OF A BINOMIAL DISTRIBUTION

The variance of a binomial distribution is $\sigma^2 = n \cdot p \cdot q$. The standard deviation of a binomial distribution is $\sigma = \sqrt{n \cdot p \cdot q}$.

We will prove the formula for expected value, but will leave the proof of the variance formula as an exercise.

Since expected value equals $\sum_{i=0}^{m} p(E_i) \cdot E_i$, and $p(E_i) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$ in a Binomial Distribution, we are interested in computing $\sum_{i=0}^{m} k \cdot \binom{n}{k} \cdot p^k \cdot q^{n-k}$. Writing out a few terms to get the feel for how this looks yields:

$$0 \cdot \binom{n}{0} \cdot p^0 \cdot q^n + 1 \cdot \binom{n}{1} \cdot p^1 \cdot q^{n-1} + 2 \cdot \binom{n}{2} \cdot p^2 \cdot q^{n-2} + 3 \cdot \binom{n}{3} \cdot p^3 \cdot q^{n-3} + \ldots + n \cdot \binom{n}{n} \cdot p^n \cdot q^0$$

The first term will clearly equal 0, so we remove it, giving us:

$$1 \cdot \binom{n}{1} \cdot p^1 \cdot q^{n-1} + 2 \cdot \binom{n}{2} \cdot p^2 \cdot q^{n-2} + 3 \cdot \binom{n}{3} \cdot p^3 \cdot q^{n-3} + \ldots + n \cdot \binom{n}{n} \cdot p^n \cdot q^0$$

Each of these terms contains a factor of p, which seems helpful. Factoring p out yields:

$$p \cdot \left[1 \cdot \binom{n}{1} \cdot p^0 \cdot q^{n-1} + 2 \cdot \binom{n}{2} \cdot p^1 \cdot q^{n-2} + 3 \cdot \binom{n}{3} \cdot p^2 \cdot q^{n-3} + \ldots + n \cdot \binom{n}{n} \cdot p^{n-1} \cdot q^0 \right]$$

We know we are looking for $n \cdot p$, so we are looking for a factor of n as well. There is a factor of n inside each $\binom{n}{k}$, since $\binom{n}{k} = \frac{n!}{(n-k)!k!}$, and so we can rewrite this as $\frac{n \cdot (n-1)!}{(n-k)!k!}$ in order to factor out the n:

$$np \cdot \left[1 \cdot \frac{(n-1)!}{(n-1)! \cdot 1!} \cdot p^0 \cdot q^{n-1} + 2 \cdot \frac{(n-1)!}{(n-2)! \cdot 2!} \cdot p^1 \cdot q^{n-2} + 3 \cdot \frac{(n-1)!}{(n-3)! \cdot 3!} \cdot p^2 \cdot q^{n-3} + \ldots + n \cdot \frac{(n-1)!}{0! \cdot n!} \cdot p^{n-1} \cdot q^0 \right]$$

Now that the combinations are broken up, it seems we can reduce each fraction a bit more. For example, $3 \cdot \frac{(n-1)!}{(n-3)! \cdot 3!}$ has a factor of 3 in the numerator and denominator, so this can be rewritten as $\frac{(n-1)!}{(n-3)! \cdot 2!}$. In general, $k \cdot \frac{(n-1)!}{(n-k)! \cdot k!}$ can be rewritten as $\frac{(n-1)!}{(n-k)!(k-1)!}$, and so now we have:

$$np \cdot \left[\frac{(n-1)!}{(n-1)! \cdot 0!} \cdot p^0 \cdot q^{n-1} + \frac{(n-1)!}{(n-2)! \cdot 1!} \cdot p^1 \cdot q^{n-2} + \frac{(n-1)!}{(n-3)! \cdot 2!} \cdot p^2 \cdot q^{n-3} + \ldots + \frac{(n-1)!}{0! \cdot (n-1)!} \cdot p^{n-1} \cdot q^0 \right]$$

What is left inside the brackets looks suspiciously like a Binomial Distribution: powers of p count up, powers of q count down, and these are multiplied by strange factorial coefficients. If we could write these coefficients as combinations, then we really would be getting somewhere. The first coefficient, $\frac{(n-1)!}{(n-1)! \cdot 0!}$, is $\binom{n-1}{0}$, and the last coefficient is $\binom{n-1}{n-1}$, so maybe these can all be written as $n-1$ choose something?

Our general coefficient is $\frac{(n-1)!}{(n-k)!(k-1)!}$ and if this is going to be written as $n-1$ choose something, it looks like that something needs to be $k-1$. Does this work? $\binom{n-1}{k-1} = \frac{(n-1)!}{[(n-1)-(k-1)]! \cdot (k-1)!}$, and since $(n-1)-(k-1) = n-k$, $\binom{n-1}{k-1} = \frac{(n-1)!}{(n-k)! \cdot (k-1)!}$. Success! We can now rewrite our main line of work as:

$$np \cdot \left[\binom{n-1}{0} \cdot p^0 \cdot q^{n-1} + \binom{n-1}{1} \cdot p^1 \cdot q^{n-2} + \binom{n-1}{2} \cdot p^2 \cdot q^{n-3} + \ldots + \binom{n-1}{n-1} \cdot p^{n-1} \cdot q^0 \right]$$

Now this can more clearly be rewritten as a Binomial Distribution: $np \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot q^{n-1-k}$

And by the Binomial Expansion Theorem, $\sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot q^{n-1-k} = (p+q)^{n-1}$. But this is a Binomial Distribution, so $p + q = 1$, and $\sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot q^{\,n-1-k} = 1$. Therefore,

$$\sum_{k=0}^{n} k \cdot \binom{n}{k} \cdot p^k \cdot q^{\,n-k}$$
$$= np \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} \cdot p^k \cdot q^{\,n-1-k}$$
$$= np \cdot (p+q)^{n-1} = np \cdot 1^{n-1} = n \cdot p, \text{ as desired!}$$

We also note the symmetry of the probabilities for the number of heads from 0 to 8. The most likely outcome is 4 although this does not occur as frequently as we might think, at "only" 27.347% of the time. The outcomes of 3 and 5 are next most likely, followed by 2 and 6, 1 and 7, and 0 and 8 being fairly rare, as we might anticipate. This type of symmetry causes the mean, median, and mode of the number of heads to be the same value. This is an important characteristic of the most important distribution in statistics: the Normal Distribution.

3.6 THE NORMAL DISTRIBUTION

Many people have heard of the bell curve, the all-important graph in statistics. This is the graph of the Normal Distribution. The Normal Distribution is symmetric, with equal mean, median, and mode. The Normal Distribution is also a probability distribution, and therefore the sum of all the probabilities in the Normal Distribution is 1. There is one important difference between the Normal Distribution and the probability distributions we have been considering thus far, however. The probability distributions we have been looking at only have a finite number of outcomes, and each occurrence within that probability distribution fits nicely into one of the different possible outcomes. These types of probability distributions are called **discrete** because each outcome is a different possibility. For example, the Binomial Distribution is a discrete probability distribution since when we flip 8 coins we can get 0 though 8 heads, but not 4.5 heads.

The Normal Distribution, on the other hand, is a **continuous** probability distribution, which means that when we use the Normal Distribution, we are considering the possibility of any outcome. The Normal Distribution is most often used to model naturally occurring data, such as the heights of adults. These heights are continuous since heights of, say, 70.23 inches are possible and in theory have no upper or lower bounds.

The fact that the Normal Distribution is continuous means calculating exact probabilities requires so phisticated calculus at the entry level and becomes extremely difficult and mathematically dicey at the more difficult levels. We will leave these calculations to upper level undergraduate (or possibility low-level graduate) courses in statistics and encourage the interested readers who are well versed in calculus to take up this line of thought. In this *Mathematics Resource Guide*, we will take a more concrete track of using the Binomial Distribution to approximate the Normal Distribution. This means we won't be able to find the exact probabilities we are looking for, but thanks to the advent of computing technology, we'll be able to get pretty close (close enough for statistics, anyway!).

But what probabilities are we even trying to calculate? The Normal Distribution models naturally occurring data, like IQ scores or heights of trees or weights of animals. What are we trying to calculate the probability of across all of these scenarios? We need a way to standardize our question, so we can apply it across a variety of contexts and situations. This is where the concept of the z-score finally comes in handy.

One of the big payoffs for the Normal Distribution is that it allows us to make the connection between z-scores and probabilities. Recall that a z-score measures how many standard deviations a particular data value is above or below the mean. If we can use the Normal Distribution to determine the likelihood of a particular data value being a certain number of standard deviations above or below the mean, we can tell how likely it is that a person has a particular IQ score or the likelihood of finding a tree that is 25 feet high or a pig that weighs 850 pounds *without* testing every person for their IQ or measuring the height of every tree or weighing every pig. The Normal Distribution, coupled with the concept of z-score, can allow us to bypass our initial definition of probability, $p(E) = \frac{\textit{number of ways event E can occur}}{\textit{total number of outcomes}}$. Although this is a wonderful mathematical definition for probability, it is very awkward and cumbersome to use in a real-life scenario, where finding the total number of outcomes may be extremely difficult or impossible.

So how can we determine the probability of having a particular z-score? We will use the Binomial Distribution with $p = 0.5$ to approximate the Normal Distribution by letting n become an extremely large number. (Note that this is *not* how statisticians calculate these values—they use calculus.) Let's say we want to know the probability of being within one standard deviation of the mean, so a z-score between –1 and 1, inclusive. Since we know the mean (expected value) of the Binomial Distribution is $n \cdot p$ and the standard deviation is $\sqrt{n \cdot p \cdot q}$, we can find the probability of a result within one standard deviation of the mean for subsequently larger and larger values of n. This will eventually approximate the probability of being within one standard deviation of the mean under the Normal Distribution.

We will select values of n that result in whole numbers for $n \cdot p$ and $\sqrt{n \cdot p \cdot q}$ in order to make it easier to locate the boundaries of being within one standard deviation of the mean. The first such value is $n = 16$, which since $p = 0.5$ gives an expected value of 8 and a standard deviation of 2. Therefore, we are interested in the probability of getting between 6 and 10 successes out of 16, using a Binomial Distribution with

$p = q = 0.5$. The probability of 6 out of 16 successes is $\binom{16}{6} \cdot (0.5)^6 \cdot (0.5)^{10}$, or $\binom{16}{6} \cdot (0.5)^{16}$. The probability of 7 successes is $\binom{16}{7} \cdot (0.5)^{16}$, the probability of 8 successes is $\binom{16}{8} \cdot (0.5)^{16}$, 9 successes is $\binom{16}{9} \cdot (0.5)^{16}$, and 10 successes is $\binom{16}{10} \cdot (0.5)^{16}$. We would like to sum these probabilities, and therefore we write this calculation in sigma notation as $\sum_{i=6}^{10} \binom{16}{i} \cdot (0.5)^{16}$. This has an approximate value of 0.7898864. As the value of n is allowed to increase, the value of this summation should approach the probability of being within one standard deviation of the mean for the Normal Distribution. The results of several values of n are shown in the table below.

n	$\overline{E}$	σ	$\overline{E} \pm \sigma$	$\sum$	APPROXIMATE PROBABILITY
16	8	2	6 – 10	$\sum_{i=6}^{10} \binom{16}{i} \cdot (0.5)^{16}$	0.7898864
64	32	4	28 – 36	$\sum_{i=28}^{36} \binom{64}{i} \cdot (0.5)^{64}$	0.7395642
256	128	8	120 – 136	$\sum_{i=120}^{136} \binom{256}{i} \cdot (0.5)^{256}$	0.7120107
1024	512	16	496 – 528	$\sum_{i=496}^{528} \binom{1024}{i} \cdot (0.5)^{1024}$	0.6975788
4096	2048	32	2016 – 2080	$\sum_{i=2016}^{2080} \binom{4095}{i} \cdot (0.5)^{4096}$	0.6901923
16384	8192	64	8128 – 8256	$\sum_{i=8128}^{8256} \binom{16384}{i} \cdot (0.5)^{16384}$	0.6864555
65536	32768	128	32640 – 32896	$\sum_{i=32640}^{32896} \binom{65536}{i} \cdot (0.5)^{65536}$	0.6845620
262144	131072	256	130816 – 131328	$\sum_{i=130816}^{131328} \binom{262144}{i} \cdot (0.5)^{262144}$	0.6836337

As n increases, the change in the probability of being within one standard deviation of the mean is decreasing as well, so it seems at $n = 262{,}144$ we are fairly close to the actual value of being within one standard deviation of the mean. Therefore, we will say that approximately 68% of data values lie within one standard deviation of the mean under the Normal Distribution.

What about the percentage that lies within two standard deviations of the mean, or three standard deviations of the mean? Using $n = 262{,}144$ again, we can approximate these values as well. Since $\overline{E} = 131072$ and $\sigma = 256$, to be within two standard deviations, we need to be between 130,560 and 131,584 successes.

$\sum_{i=130560}^{131584} \binom{262144}{i} \cdot (0.5)^{262144} \approx 0.9547104$, so we will say approximately 95% of data values fall within two standard deviations of the mean under the Normal Distribution. For three standard deviations, we need to be between 130,304 and 131,840 successes, and as $\sum_{i=130304}^{131840} \binom{262144}{i} \cdot (0.5)^{262144} \approx 0.9973002$, we will say approximately 99.7% of data values fall within three standard deviations of the mean under the Normal Distribution.

PROBABILITIES FOR THE NORMAL DISTRIBUTION

Under a Normal Distribution, approximately 68% of data values fall within one standard deviation of the mean, approximately 95% of data values fall within two standard deviations of the mean, and approximately 99.7% of data values fall within three standard deviations of the mean.

These percentages for the Normal Distribution are sometimes referred to as the **Empirical Rule**.

These percentages only hold for a Normal Distribution, and—like hidden assumptions of independent probabilities—data sets are often assumed to be normally distributed when they may not be normally distributed. When dealing with data collected from an experiment, use caution when assuming the data are normally distributed.

These probabilities, together with the idea that in a Normal Distribution half of the data values fall above the mean and half fall below the mean, allow us to answer the types of questions raised at the end of our discussion of z-scores earlier in the resource guide. Recall that a z-score measures the number of standard deviations a data value is above or below the mean. Therefore, we can restate the probabilities for the Normal Distribution in terms of z-scores.

PROBABILITIES FOR THE NORMAL DISTRIBUTION IN TERMS OF Z - SCORES

For a data set that is normally distributed, approximately 68% of the data values will have z-scores between –1 and 1 (inclusive), approximately 95% of the data values will have z-scores between –2 and 2 (inclusive), and approximately 99.7% of the data values will have z-scores between –3 and 3 (inclusive).

EXAMPLE 3.6A: Assume the average height of people in a major metropolitan area is 68.5 inches, with a standard deviation of 2.5 inches. What approximate percentage of the population is between 66 and 71 inches tall?

SOLUTION: The z-score for 66 inches is –1, and the z-score for 71 inches is 1, so approximately **68%** of the population is between 66 and 71 inches tall.

EXAMPLE 3.6B: In the same metropolitan area referred to in Example 3.6a, what is the probability of meeting someone who is 76 inches or taller?

SOLUTION: The z-score for 76 inches using this mean and standard deviation is +3. We know that 99.7% of the population should have z-scores between –3 and +3, so 0.3% of the population falls outside this range. But this 0.3% of the population has z-scores less than –3 or greater than +3, and we are interested only in those greater than +3. Therefore, the probability of meeting someone who is taller than 76 inches is **0.15%**, half of 0.3%.

EXAMPLE 3.6C: In this same metropolitan area, how tall does someone need to be so that at least 2.5% of the population is shorter than they are?

SOLUTION: Since 95% of z-scores are within 2 standard deviations of the mean, the bottom 2.5% are people whose z-score is less than –2. A z-score of –2 corresponds to a height of 63.5 inches, so a person who is **63.5 inches or taller** will have at least 2.5% of the population shorter than they are.

We can hopefully see from these examples that there is a correspondence between a percentage and a z-score under a Normal Distribution. With the information presented here, we are only aware of this correspondence for a few integer-value z-scores. Most z-scores, however, are not going to be nice integer values like –3 or +2, so it seems a more general correspondence between z-scores like 1.4 or –2.37 and percentages would be helpful. Indeed, this development is at the heart of undergraduate statistics, and it is not our purpose here to explore this idea, but merely to point out the next logical step in this progression.

As stated earlier, many naturally occurring variables do follow the Normal Distribution, but many data sets will not be normal. Indeed, most data collected are not from populations that follow the Normal Distribution. So why is the Normal Distribution so important? It turns out that even if the original set of data is not normally distributed, when small samples are taken from a data set and the distribution of sample means is considered, this distribution of sample means can be modeled by the Normal Distribution. This important idea is called the Central Limit Theorem. Again, a discussion of this theorem is beyond the scope of this guide, but we mention it here to give the reader some foresight into undergraduate statistics.

SECTION 3 SUMMARY: STATISTICS

- **Measures of Central Tendency:** The three measures of central tendency of a set of data are the **mean**, **median**, and **mode**.
- **Definition and Formula for the Mean:** The **mean** of a data set is the value each data point would have if all data points were the same value. The mean of a set consisting of n values is given by $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, where x_i represents the i^{th} data value.
- **Definition and Formula for the Median:** The **median** of a data set is the value with the same number of data points above as below the value. If n is odd, the median is $x_{\frac{n+1}{2}}$; if n is even, the median is the mean of the two data values at the middle of the data: $\frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$.
- **Definition of the Mode:** The **mode** is the data value that occurs the most frequently. If each value in the data set occurs once, the data have no mode. If two different data values each occur most frequently, they are both considered the mode, and the data set is called **bimodal**.
- **Mean and Median of a Finite Arithmetic Sequence:** The mean and the median of a finite arithmetic sequence are always equal.
- **Measures of Spread:** The measures of the spread of a set of data are **range**, **interquartile range**, **variance**, and **standard deviation**.
- **Definition of Range:** The **range** of a data set is the difference between the highest and lowest value.
- **Definition of Q_3 and Q_1:** The **lower quartile** or **first quartile** (abbreviated Q_1) is the median value of the data below the median in a set. The **upper quartile** or **third quartile** (abbreviated Q_3) is the median value of the data above the median in a set.

- **Definition of Interquartile Range:** The **interquartile range** (IQR) of a data set is the difference between the **third quartile** (Q_3) and **first quartile** (Q_1).

- **IQR Test for Outliers:** A data point x_i is considered an outlier if $x_i > Q_3 + 1.5 \cdot \text{IQR}$ or $x_i < Q_1 - 1.5 \cdot \text{IQR}$.

- **Definition and Formulas for Variance:** The **variance** of a set of data is the average of the squared difference of each data point from the mean. Variance is denoted by the Greek letter sigma squared, written as σ^2. In symbols, $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$. There are alternative formulas for the variance, including $\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - \bar{x} \cdot \sum_{i=1}^{n} x_i}{n}$ and $\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n}$.

- **Definition and Formula for the Standard Deviation:** The standard deviation of a set of data, denoted by the Greek letter sigma, is the square root of the variance for the data: $\sigma = \sqrt{\sigma^2}$. The rationale for taking the square root of the variance is to have a statistical measure with the same units as the original data.

- **Definition and Formula for the Z-Score:** A **z-score** represents the number of standard deviations a particular data value is above or below the mean. A z-score is calculated by the formula $z = \frac{x_i - \bar{x}}{\sigma}$.

- **Definition and Formula for Probability:** The **probability** of an event occurring is the number of ways that event can occur divided by the total number of possible outcomes. The probability of an event E occurring is given by $p(E)$, where $p(E) = \frac{\textit{number of ways event E can occur}}{\textit{total number of outcomes}}$.

- **Basic Properties of Probabilities:** There are three basic properties of probabilities.
  - The probability of any event is a number between 0 and 1. An event with probability 0 is impossible, and an event with probability 1 is certain. In symbols, for any event E, $0 \le p(E) \le 1$.
  - The probability of an event occurring plus the probability of the event not occurring is 1. The probability of an event E not occurring is called the **complement** of E and is denoted by E'. In symbols, $p(E) + p(E') = 1$.
  - In a given situation, the sum of the probabilities of all possible events must equal 1, since one of the possible events must occur. In symbols, if there are n different possible outcomes, $\sum_{i=1}^{n} p(E_i) = 1$.

- **Definition of Independent Events:** Two (or more) events are called **independent** if the occurrence of one event does not affect the occurrence of the other event(s). When determining the probabilities of each event, the events can be considered separately since there is no meaningful interaction between the events.

- **Probability of Independent Events:** The probability of two (or more) independent events occurring together is the product of the probabilities of each individual event. If there are two events A and B, this is written as $p(A \text{ and } B) = p(A) \cdot p(B)$. If there are n events, this is written as $p(E_1 \text{ and } E_2 \text{ and} \ldots \text{ and } E_n) = p(E_1) \cdot p(E_2) \ldots p(E_n)$.

- **Definition of Dependent Events:** Two (or more) events are called **dependent** if the occurrence of one event does affect the occurrence of the other event(s). When determining the probabilities of each event, the events cannot be considered separately, as the events are connected in a meaningful way.

- **Probability for Dependent Events:** The probability of two dependent events occurring together is the product of the probability of the first event occurring and the probability of the second event occurring given that the first event occurred. In symbols, if there are two events A and B, this is written as $p(A \text{ and } B) = p(A) \cdot p(B|A)$. $p(A)$ represents the probability of event A occurring, and $p(B|A)$, read aloud as "the probability of B given A," represents the probability that event B occurs if event A has already occurred.

- **Definition of a Probability Distribution:** A probability distribution is a set of outcomes where the probability of each event is a number between 0 and 1, and the sum of the probabilities from all outcomes equals 1. Symbolically, if there are m different outcomes, for all events E_i within the probability distribution, $0 \leq p(E_i) \leq 1$ and $\sum_{i=1}^{m} p(E_i) = 1$.

- **Definition and Formula for Expected Value:** The expected value of a probability distribution is the mean payoff expected from the probability distribution over a long enough period of time. A probability distribution with m different outcomes $E_1, E_2, \ldots E_m$, each with a corresponding probability of occurrence $p(E_1), p(E_2), \ldots p(E_m)$, has expected value $\overline{E} = \sum_{i=1}^{m} p(E_i) \cdot E_i$.

- **Definition of Fair Game:** A **fair game** is a probability distribution with an expected value of 0.

- **Variance of a Probability Distribution:** The variance of a probability distribution is the sum of the squares of the differences between each outcome and the expected value multiplied by the probability of each outcome: $\sigma^2 = \sum_{i=1}^{m} (E_i - \overline{E})^2 \cdot p(E_i)$. The alternative formula for the variance of a probability distribution is $\sigma^2 = \sum_{i=1}^{m} [E_i^2 \cdot p(E_i)] - \overline{E}^2$.

- **Standard Deviation of a Probability Distribution:** The standard deviation of a probability distribution is the square root of the variance of the probability distribution: $\sigma = \sqrt{\sum_{i=1}^{m} (E_i - \overline{E})^2 \cdot p(E_i)}$ or $\sqrt{\sum_{i=1}^{m} [E_i^2 \cdot p(E_i)] - \overline{E}^2}$.

- **Criteria and Formula for the Binomial Distribution:** A situation can be modeled using the Binomial Distribution if the following criteria hold: each outcome is made up of a series of independent events, these events can be classified as success or failure, and the probability of success is constant across events. Given a probability of success p and a probability of failure q for each event, the probability of k successes out of n events is given by the formula $\binom{n}{k} \cdot p^k \cdot q^{n-k}$. Since p is the probability of success and q is the probability of failure, $p + q = 1$. With n events, there are $n + 1$ different possible outcomes since it is possible for the event to occur 0 out of n times.

- The Binomial Distribution is a probability distribution.

- **Expected Value for a Binomial Distribution:** The expected value of a binomial distribution with n events and probability of success p for each event is $\overline{E} = n \cdot p$.

- **Variance and Standard Deviation of a Binomial Distribution:** The variance of a binomial distribution is $\sigma^2 = n \cdot p \cdot q$. The standard deviation of a binomial distribution is $\sigma = \sqrt{n \cdot p \cdot q}$.

- **Approximation of the Normal Distribution:** We can use the Binomial Distribution with $p = 0.5$ and increasing values of n to approximate the Normal Distribution.

- **Probabilities for the Normal Distribution:** Under a Normal Distribution, approximately 68% of data values fall within one standard deviation of the mean, approximately 95% of data values fall within two standard deviations of the mean, and approximately 99.7% of data values fall within three standard deviations of the mean.

- **Probabilities for the Normal Distribution in Terms of Z-Scores:** For a data set that is normally distributed, approximately 68% of the data values will have z-scores between –1 and 1 (inclusive), approximately 95% of the data values will have z-scores between –2 and 2 (inclusive), and approximately 99.7% of the data values will have z-scores between –3 and 3 (inclusive).

SECTION 3 REVIEW PROBLEMS: STATISTICS

1. Consider the following set of data: 7, 8, 13, 9, 10, 11, 5, 4, 3, 4, 5, 6, 5, 5, 3, 4.

    a. Find the mean, median, and mode of these data.

    b. Find the range and interquartile range for these data.

    c. Find the variance and standard deviation for these data.

    d. What is the z-score for the data value 4 in the above data?

    e. Are any of these data outliers? Explain.

2. Some parents are wondering if there is a difference in the amount of homework students are assigned at two local high schools: Rydell High School and Shermer High School. To try to test this, twenty students are selected at random from each high school and asked to record how many hours they spend doing homework during the next week. The results are shown below:

RYDELL HIGH SCHOOL	13	15	7	9	5	16	6	5	11	14	15	8	8	6	10	12	9	8	13	5
SHERMER HIGH SCHOOL	12	11	3	7	6	10	2	5	12	4	8	6	5	9	5	8	4	8	7	11

    a. Find the mean, median, and mode of the Rydell High data.

    b. Find the mean, median, and mode of the Shermer High data.

    c. Find the range and interquartile range of the Rydell High data.

    d. Find the range and interquartile range of the Shermer High data.

    e. Find the variance and standard deviation of the Rydell High data.

f. Find the variance and standard deviation of the Shermer High data.

g. Which school do you believe assigns more homework? Explain and justify your answer.

3. Consider the following breakdown of 1^{st} place, 2^{nd} place, and 3^{rd} place from a recent track season for Rydell High School, Ridgemont High School, Bayside High School, and Shermer High School. A total of 100 events were run during the season.

	1^{st}	2^{nd}	3^{rd}
RYDELL	36	19	28
RIDGEMONT	24	27	37
BAYSIDE	26	18	22
SHERMER	14	36	13

a. If a top-three finisher is selected at random, what is the probability he/she is from Rydell High-School?

b. If a top-three finisher is selected at random, what is the probability he/she is from Bayside and finished 3^{rd}?

c. Given that a 1^{st} place finisher is selected, what is the probability he/she is from Ridgemont High School?

d. Given that a top-three finisher from Ridgemont High School is selected, what is the probability he/she finished 1^{st}?

e. If a top-three finisher is selected, what is the probability he/she finished 3^{rd}?

f. If a top-three finisher is selected, what is the probability he/she is from Shermer High School?

g. If a top-three finisher is selected, what is the probability he/she is from Shermer High School and finished 3^{rd}?

h. Is being a 3^{rd} place finisher and being from Shermer High School independent of each other? Explain.

4. Consider the following probability distribution.

OUTCOME	5	6	7	8	9
PROBABILITY	0.2	0.25		0.1	0.07

a. What value is needed to complete the probability distribution? Explain.

b. What is the expected value of this probability distribution?

c. What are the variance and standard deviation of this probability distribution?

5. A fair twenty-sided die numbered 1–20 is rolled. What is the probability of getting:

   a. An odd number?

   b. A number divisible by 3?

   c. A number divisible by 5?

   d. A number divisible by 3 or 5?

   e. A prime number (1 is not prime)?

   f. A number divisible by 2 and 3?

6. A set of seven (7) 20-sided dice are rolled. What is the probability of getting:

   a. Exactly three (3) results of 19?

   b. At least two (2) results of 11?

   c. At most four (4) results of 17?

7. A set of seven (7) 20-sided dice are rolled.

   a. What is the expected value for the number of 9's that are rolled?

   b. What are the standard deviation and the variance for the number of 9's that are rolled?

8. Consider a standard deck of playing cards (52 cards, no jokers). If two cards are drawn without replacement, what is the probability that the two cards drawn are:

   a. Both spades?

   b. Both kings?

   c. Two mismatched cards (do not share suit or face value)?

   d. Both queens?

e. A pair (same face value)?

f. Two matching cards (have either the same suit or the same face value)?

9. If five cards are drawn from a standard deck without replacement, what is the probability of drawing:

   a. Exactly three kings?

   b. Any three of a kind (three cards with the same value, two cards with different values)?

   c. A full house (a three of a kind and a pair?)

10. Francine likes to watch for shooting stars in the night sky. In any given hour, the probability that Francine sees a shooting star is 45%. Francine watches the night sky for shooting stars for two consecutive hours.

    a. What is the probability she sees exactly two shooting stars?

    b. What is the probability she sees exactly one shooting star?

    c. What is the probability she sees no shooting stars?

    d. What is the probability she sees a shooting star in the second hour, given that she saw one in the first hour?

    e. Construct a probability distribution to model this situation.

    f. Find the mean, variance, and standard deviation for the number of shooting stars Francine will see.

11. The probability that a bat will fly into a house is 0.02. The probability that a bat will have rabies and fly into a house is 0.0096.

    a. Assuming the events of a bat flying into a house and a bat having rabies are independent, what is the probability of a bat having rabies?

    b. Do you believe that a bat flying into a house is independent of a bat having rabies?

    c. Tests on bats captured in the wild suggest that the probability of a randomly selected bat having rabies is 0.05. What does this imply about the assumed independence of a bat flying into a house and a bat having rabies?

d. What is the probability that a bat has rabies, given that it has flown into a house?

12. On average, a family of four in the United States spends $150 per week on food with a standard deviation of $7.50. Assume these data are normally distributed.

    a. What is the z-score of a family that spends $140 per week on food?

    b. What is the z-score of a family that spends $185 per week on food?

    c. The middle 68% of families spends between what amounts on food?

    d. The middle 95% of families spends between what amounts on food?

    e. What is the probability of selecting a family that spends less than $135 per week on food?

    f. What is the probability of selecting two families that both spend less than $135 per week on food?

    g. What is the probability of selecting three families that all spend less than $135 per week on food?

    h. If three families were selected that all spent less than $135 per week on food, would you think this was a random selection? Explain.

    i. If three families were randomly selected, and they all spent less than $135 on food, what would you think about this situation? Explain.

13. An economist wants to study the average income of families in a local metropolitan area. She randomly selected 55 anonymous tax returns from the IRS, and the results are shown below.

$35,861	42,403	47,601	13,519	54,190
13,730	59,732	10,014	50,757	107,869
93,181	33,291	18,821	92,107	32,269
58,079	48,145	75,601	48,351	58,641
28,992	76,972	34,811	65,919	9,710
48,633	33,711	79,708	42,381	56,951
44,179	80,531	40,079	77,460	10,028
28,456	22,105	80,197	23,986	54,507
91,097	57,706	33,468	38,131	77,349
56,660	36,511	40,788	59,832	28,466
24,472	46,325	110,578	43,548	81,648

a. Find the mean, median, and mode of these data. What does a typical family in this metropolitan area earn as an annual income?

b. Find the range and interquartile range for these data.

c. Find the variance and standard deviation for these data.

d. Are there any outliers in these data? Explain.

e. What approximate incomes would be in the middle 68% of the data given?

f. What is the z-score for the income of $47,601? What does this mean?

g. Assume that income is normally distributed in this area, with mean and standard deviation equal to the values determined above. Under this assumption, what incomes represent the middle 68% of annual income in this metropolitan area?

h. What do your answers to parts "e" and "g" imply when compared to each other? Explain.

14. In order to try to increase attendance, a local casino has developed a new card game to lure gamblers, called *The Joke's on You!* The game is advertised as being very straightforward; to play, gamblers just draw a single card out of a deck of 54 cards (standard deck including two jokers). The payout scheme is as follows:

IF A GAMBLER DRAWS	THE PAYOUT IS
the ace of spades	$25
an ace (not of spades)	$10
the king of hearts	$9
a king or a heart (but not K of ♥)	$8
anything else	$0
a joker	–$20

None of these categories are allowed to overlap. Each card fits into only one category, and it is always the category that is best for the gambler.

The cost to play the game is $2. If the gambler draws a joker, he/she must pay the dealer $20 more (from his/her pocket).

a. Calculate the probabilities of each of the six outcomes listed above on a single play of *The Joke's on You!*

b. Construct a probability distribution using these probabilities and the outcome for the player. (Don't forget it costs \$2 to play.)

c. Verify that your probability distribution is a probability distribution. Explain how you know it is a probability distribution.

d. Calculate the expected value, variance, and standard deviation for this probability distribution.

e. Would you recommend that gamblers play this game? Why or why not?
Would you recommend that the casino offer this game? Why or why not?

f. Explain why this situation *cannot* be modeled using the binomial theorem/distribution.

15. Prove that the variance for the Binomial Distribution is $\sigma^2 = n \cdot p \cdot q$.

Conclusion

This year, the *Mathematics Resource Guide* focused on permutations and combinations, algebra, and statistics. Our goal with respect to these topics was twofold: to highlight important areas of mathematics that are often deemphasized or omitted altogether from a traditional high school mathematics sequence and to investigate important and interesting connections between these apparently disparate areas of mathematics.

In Section I, we looked at the mathematical ideas of counting arrangements and groups. Permutations and combinations are very flexible and important mathematical ideas and crop up in all sorts of unexpected places. Combinations are particularly useful, as we saw them appear several times in our discussion of algebra and statistics later in the resource guide.

In Section II , which focused on the topic of algebra, we discussed portions of algebra not normally addressed in high school mathematics. The general impression of algebra that most people have from studying it in high school is a potentially distorted, limited picture of what mathematicians consider when they think about algebra, and our goal here was to give you a broader sense of what algebra is. Arithmetic and geometric sequences and series, sigma notation, and polynomials are all important algebraic structures that deserve our careful attention. We studied the Binomial Expansion Theorem, an incredibly important and foundational result that is often left out of high school mathematics or studied without giving students a true sense of its power and importance. We looked at several applications of the Binomial Expansion Theorem, most importantly in statistics with the Binomial Distribution. With our study of Euler's constant, we saw an interesting example of mathematics generating more mathematics, which is truly what undergraduate and graduate level mathematics are based upon.

In the third and final section, covering statistics, we studied measures with which you were probably already familiar, such as mean, median, mode, range, variance, and standard deviation, but we tried to illuminate the reasons for these measures and show that the formulas for these measures were developed in a manner that makes sense. We also studied the foundational topic of probability distributions and looked at two important distributions: the Binomial Distribution and the Normal Distribution. Through proving important results about the Binomial Distribution, you have hopefully gained an appreciation for the mathematical foundations of statistics as well as the utility of sigma notation and the Binomial Expansion Theorem.

Our goal was to make this year's *Mathematics Resource Guide* challenging, enlightening, and interesting. Although sometimes misrepresented as a sequence of unconnected topics, mathematics at its core is about seeking and finding interesting connections between ideas, and we hope you are left wondering about other connections within and between mathematics that you may have previously overlooked.