P. L. Moore

Quantitative Problem Solving in Natural Resources

OCTOBER 5, 2018

Iowa State University Digital Press

A GREAT DISCOVERY SOLVES A GREAT PROBLEM BUT THERE IS A GRAIN OF DISCOVERY IN THE SOLUTION OF ANY PROBLEM. YOUR PROBLEM MAY BE MODEST; BUT IF IT CHALLENGES YOUR CURIOSITY AND BRINGS INTO PLAY YOUR INVENTIVE FACULTIES, AND IF YOU SOLVE IT BY YOUR OWN MEANS, YOU MAY EXPERIENCE THE TENSION AND ENJOY THE TRIUMPH OF DISCOVERY. SUCH EXPERIENCES AT A SUSCEPTIBLE AGE MAY CREATE A TASTE FOR MENTAL WORK AND LEAVE THEIR IMPRINT ON MIND AND CHARACTER FOR A LIFETIME.

GEORGE PÓLYA, *HOW TO SOLVE IT*

PUBLISHED BY IOWA STATE UNIVERSITY DIGITAL PRESS

TYPESET WITH THE TUFTE-LATEX CLASS: TUFTE-LATEX.GITHUB.IO/TUFTE-LATEX/

Contents

Preface

This book is a collection of resources for students of natural resource science and management[1]. Many of you will be required to complete some mathematics and/or statistics courses during your undergraduate studies, and some of you will likely dread the prospect.

[1] I interpret this broadly, so that we may include study of fisheries, wildlife, forestry, and water resources, among other things

However, if you're reading this, you have probably been presented with an alternative means of satisfying at least some of your quantitative skills requirements. This is the origin of the course NREM 240 at Iowa State University, for which this book was originally prepared. NREM 240 is a class about solving quantitative problems that non-mathematicians interested in biology and environmental sciences may find compelling.

I use the word *problem* here in the same way that math education scholar Alan Schoenfeld does — to describe an intellectual challenge that is quantitative in nature[2]. A problem is distinct from an exercise in subtle, but crucial ways. An exercise is a prompt for which a student must select one or more of a small number of recently-demonstrated procedures or algorithms to reveal a clear and known[3] result. Contrast that with the intellectual impasse of Schoenfeld. In this spirit, a problem is a deeper challenge that often contains a mixture of complexity, uncertainty, and ambiguity and requires some technical skill or knowledge. A problem also often requires creativity — or as Pólya says in the quote[4], "inventive faculties" – and the willingness to explore, to try and fail, to persist, and to learn from mistakes. This kind of task can be frustrating and uncomfortable for those of us not accustomed to it, particularly if we have no interest or investment in the problem being posed. But life rarely provides us with exercises. If we *are* confronted with exercises, we become bored and uninspired by the idle redundancy[5]. But a completely different sense of achievement and satisfaction is realized when we solve problems because, by their very nature, they lead us to new ways of thinking and understanding, if we are willing.

[2] Schoenfeld, A. H., 1985, *Mathematical Problem Solving*, Academic Press.

[3] at least to the instructor and the owner of a solutions manual

[4] Excerpted from the preface of Pólya, G., 1945, *How to Solve It*, Princeton University Press.

[5] Mathematician Paul Lockhart has written an engaging, though scathing critique of current school math curriculum in a book called *A Mathematician's Lament*.

NREM 240 at ISU was initially conceived as a bridge to link the quantitative skills developed elsewhere with some common applications in the natural sciences. Experience has shown that, once students know which techniques to apply to which given quantities, the computational task is rarely challenging. A greater challenge is the selection of appropriate techniques and assembly of relevant input quantities when neither are given. Since both of these processes are hallmarks of authentic problem solving, NREM 240 evolved to embrace the problem-solving process as the central objective, calling upon quantitative concepts and techniques as needed to address particular problems. Therefore, applied problems and the strategies used to address them have become the focus of the course, and this text has evolved to support that focus.

The philosophy of this text has been strongly influenced by John Harte's fantastic books, *Consider a Spherical Cow* and *Consider a Cylindrical Cow*[6]. Both books pose environmentally-themed problems that are compelling and maddeningly open-ended. But Harte boldly demonstrates how idealizations, approximations, and analogies can be leveraged to make sense of complex problems. On a tip from a friend in grad school, I picked up a copy of each of these books and was amazed. I had always been an average or slightly-below-average student of math all the way through grade school and college, and had never enjoyed it. Harte's problems hooked me because they offered a means to address problems I found compelling. I saw, perhaps for the first time, a way that math could help me gain insights into things I already wanted more insights into. I experienced an unfamiliar willingness to labor over unit conversion details, to really wrestle with what it meant to integrate a function, and to chase wild and risky ideas to see where they led. That feeling has stuck with me over the years, and I have sought to facilitate some semblance of that experience in the students I now teach.

[6] Harte, J., 1988, *Consider a Spherical Cow: A Course in Environmental Problem Solving*, Sausalito, CA, University Science Books; Harte, J., 2001, *Consider a Cylindrical Cow: More Adventures in Environmental Problem Solving*, Sausalito, CA, University Science Books.

The methods or strategies that are highlighted here are drawn in large part from 's problem-solving framework and derivatives thereof. Excellent tutorials on problem-solving include *Thinking Mathematically* by Mason, Burton and Stacey[7], *Crossing the River with Dogs* by Johnson, Herr, and Kysh[8], and *Ants, Bikes, and Clocks* by Briggs[9]. Notwithstanding the odd titles of some of those books, all provide interesting perspectives and tips for problem-solving. Even so, many of the problems in these texts are of the sort that you might find on standardized tests: "Janet leaves on a train heading east at 42 miles per hour, while Mark...". If there is truth to the idea that we embrace challenges when they address topics that we are already interested in, these approaches may still fail to engage students. With this in mind, the examples and exercises in this text are aimed at engaging the natural resource student in thinking about for-

[7] Mason, J., L. Burton, and K. Stacey, 2010, *Thinking Mathematically*, 2nd ed., Pearson Education Ltd.

[8] Johnson, K., T. Herr, and J. Kysh, 2012, *Crossing the River with Dogs: Problem Solving for College Students*, 2nd ed., Wiley.

[9] Briggs, W.L., 2005, *Ants, Bikes and Clocks: Problem Solving for Undergraduates*, Society for Industrial and Applied Mathematics.

est measurements, fisheries management, habitat conservation, and the like.

You will find in the pages that follow an introduction to problem solving *as a process*. You'll also find a review of some frequently-used mathmatical concepts and procedures. For some of the more powerful concepts and techniques, exercises are provided to assist with reviewing (or exploring for the first time) by doing. The booklet is not intended to be a guide to be followed through a series of skills, but rather a resource to support the problem-solving process and help lower the conceptual and computational barriers along the way.

Schoenfeld argues that what constitutes a problem to you depends strongly upon your experience and formal training. This relativity makes it challenging to keep a diverse group of students on the same page, and able to succeed at similar rates. With this in mind, in preparing the materials for this book and the course it supports, I have presumed 1) that you are sincerely interested in the natural sciences; and 2) that you have had a typical sequence of high school math courses, including a few years of algebra, some geometry, and perhaps trigonometry and statistics. It may also be true that you have found it challenging or even unpleasant when asked to recruit your quantitative skills to interpret information or address a question in the coursework in your discipline. If any of this describes you, you're in the right place. Let's go!

Note to instructors

I have written this book with my students and their needs in mind, but hope that others may find it useful. It may be worth taking a moment to clarify how I use this book. While I expect my students to read this text and work on the exercises in it, I build the course around a group of "focus problems" not contained here, but with an open-ended form and practical flavor like the sample problems in Chapter 3. As we work on the first focus problem, I guide students through the problem-solving process described in Chapter 2, making frequent reference to examples contained in the book. Therefore, to appreciate the processes involved, I think it is wise to have students read the first three chapters early. The remainder of the book is organized more by problem type, and can be assigned or referred to as appropriate to support the focus problems, rather than proceeding linearly through each chapter.

I have deliberately avoided discussing particular software tools or internet resources, partly in the interest of ensuring that this text does not rapidly become obsolete, but also to allow for flexibility. In online video tutorials and in-class exercises, I ask my students to work with data and create graphs using spreadsheets, but those seeking richer computing environments can just as easily use this text alongside instruction in R, Matlab, Mathematica, and others.

1
Introduction

1.1 Some Philosophical Notes

In a conventional math course, you might be confronted with a question like the following:

> Find the roots of x in the expression:
>
> $$4x^2 - 13x + 6 = 0.$$

You may be instructed or implicitly expected to apply an algorithm to this problem and provide the two possible roots. Most likely this would be an opportunity to use the trusty quadratic formula:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

If you weren't already turned off, you might plug the values 4, -13, and 6 in for a, b, and c in the quadratic formula, perform some arithmetic and find the roots to be 0.56 and 2.69. Alternatively, if you're like me, you'd let a computer program like Geogebra[1] apply this algorithm, since that is what computers are for.

[1] Geogebra is a free, multi-platform software package that combines a CAS (computer algebra system) with a dynamic, interactive geometry and graphing interface.

In any case, with these two roots in hand, you have an answer, isn't that satisfying!? OK, maybe it is for some of you, but this has never been satisfying for me. No meaning was ever assigned to the variables or constants, nor was it claimed that the result had any context or significance. I don't really know what to *do* with the answer, now that I have it. I've learned very little, except perhaps that I can enter the proper numbers into an algorithm. That is not to say that learning the algorithm is without value – indeed it is very valuable. But for most of us, the algorithm itself is not an end in itself, it is a means to an end. It is a useful tool that allows us a shortcut to a result when an equation presents itself in a quadratic form.

Though the problems in a conventional math class may look arbitrary, they are often designed to be "well-behaved". You wouldn't often see equations exactly like the example above, because the roots turn out to be icky decimal numbers rather than nice, clean integers. Furthermore, things get complex (literally!) if the numerical coefficients on the left-hand side of the equation are such that the term under the square-root in the quadratic formula turn out to be negative. Such a poorly-behaved case belongs to a completely different subject in the mathematics curriculum (complex analysis), and so cannot be imposed upon an unsuspecting algebra student. However, in the "real world", there is no more reason to suspect a real result than a complex one, in those rare practical instances when one needs to find the roots of a second-order polynomial. Thus, in this approach we learn a very strict set of rules applicable only to an idealized subset of problems that may or may not have any significance outside of abstract trivia.

The approach we use in this course is to encounter math and statistics *in the process* of finding solutions to real—or at least plausible—problems in the natural sciences. Sometimes these real problems are messier than those out of a textbook. Often they will be open-ended and will require multiple steps and a variety of techniques. We'll need to decide for ourselves what tools and techniques to use, according to the needs of the problem. Being creative in math classes isn't what we've been trained to do, and at times it may be uncomfortable. That's OK, we'll take our time. But no matter the problem, we'll always have *a reason* to be doing math or statistics – every quantity in an equation will have real meaning or role, and we can apply our non-quantitative knowledge and experience with these entities to help us solve our problems.

LET'S HAVE A LOOK AT the kind of problem that a natural resource manager or ecologist might care more about, a problem that we'll return to periodically in this course. Consider the Iowa DNR's estimates of the statewide pheasant population shown in the graph below. Now consider the question that is important to many hunters and game managers around the state: what should we expect pheasant populations to look like in 5 years? The black dots in the graph above are annual results from roadside pheasants surveys, the dots are connected in chronological order with black lines[2], and a blue line traces the long-term trend. In the roadside survey, DNR biologists travel 30-mile segments of rural roads on dewy late-summer mornings, count the pheasants observed in each stretch, and compile the data across the whole state[3]. The changes from year to year in this measure of pheasant population are similar to the changes in hunter harvest and are thought to be a good indication of the pheasant population as a whole.

[2] note that connecting dots like this can imply continuity between data points, which may or may not be what we wish to indicate.

[3] for more information on the pheasant population trends and the roadside survey method, consult the IDNR small game website

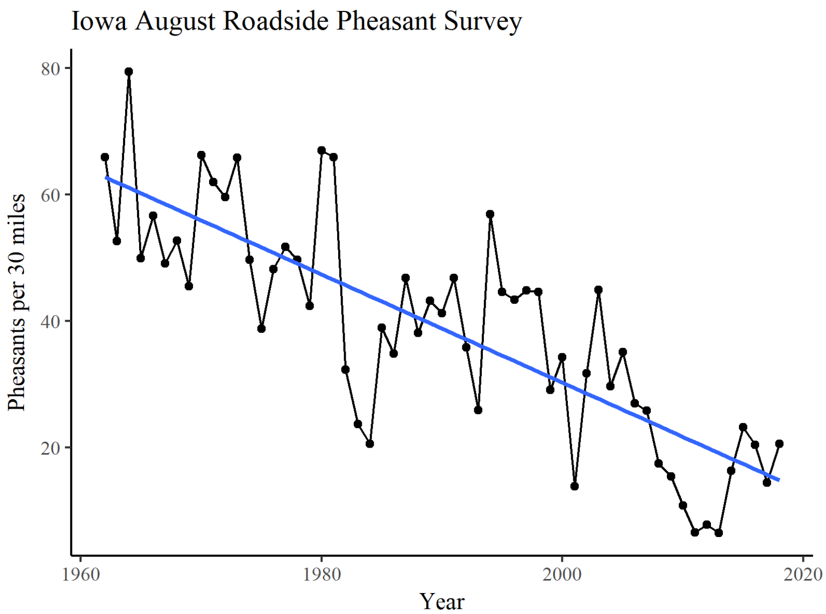


Figure 1.1: Record of pheasant counts per 30 miles from the Iowa August Roadside Pheasant Count. Data from the Iowa DNR.

The rebound in the roadside counts from 2013 to 2014 was characterized in the DNR report as a 151% increase, and the change from 2014 to 2015 was described as a 37% rise. The last few numerical value pairs plotted in the graph are also shown in Table 1.1 to the right.

year	count
2012	7.8
2013	6.5
2014	16.3
2015	23.2
2016	20.4
2017	14.4
2018	20.6

Table 1.1: The most recent five years of data from the roadside pheasant count.

This example may seem straight forward at first glance, and in some ways it is. The trend appears to indicate an overall decline in pheasant numbers across the state, and perhaps we should be prepared to take a more hands-on approach to managing pheasant numbers if we wish to sustain a viable game resource in the coming years. On the other hand, the year-to-year changes seem to be erratic, rising and falling in a way that seems to lack a pattern. Addressing the guiding question with any confidence, however, could be a bit challenging. If, for example, we were looking at the dataset at the end of 2011 following 6 straight years of steady decline, would we have been able to anticipate a rebound in 2014 or 2015? Probably not without a robust and reliable **model**[4] of the factors that cause population changes and how those factors could change in subsequent years. These are advanced topics, but ones that wildlife managers have to incorporate into their management strategies in some situations.

[4] a model in this context means an approximate mathematical representation of the real system from which predictions about the behavior of the real system may be made and tested.

On an even deeper level, the roadside pheasant count itself is a strange quantity that doesn't exactly represent what we wish to know (i.e., the pheasant population). Instead, it is an easy-to-estimate approximation of the real population. As such, it is a **sample** from the larger population, at least one step removed from the quantity we seek. How does such a sample relate to the larger quantity we are after? That's a pretty simple question in theory (i.e., in stats class), but when we account for the sampling methodology, timing, and observer variability, and we consider that pheasant visibility may not always be directly linked to population, it isn't so straight-forward after all.

The observations we've made from this dataset are just some of the many complexities that we might uncover as we endeavor to solve problems in pheasant population or habitat management in Iowa. This example hints at the concepts of time series analysis, forecasting, and measurement uncertainty, as well as functional relationships between multiple variables and between samples and populations. Each of these concepts represents a quantitative tool that can be applied toward the larger problem-solving task. We'll visit most of these concepts and many more en route to addressing practical problems and methods. But as we will see in the next chapter, the quantitative procedures that we employ in the problem-solving process are just a part of the arsenal necessary to solve practical problems. Furthermore, it should go without saying that some of the quantitative tools we do have at our disposal are *not* appropriate for some problems, and a key job of the problem-solver is to determine which tools those are. We'll delve deeper into this issue a few pages ahead.

1.2 An ancient puzzler

> "In a lake the bud of a water-lily was observed, one span above the water, and when moved by the gentle breeze, it sunk in the water at two cubits' distance. Required the depth of the water."
>
> -Henry Wadsworth Longfellow

The lines above are from a poem in which a Mr. Churchill playfully challenges his wife with mathematical puzzlers. The problem he describes actually dates back many centuries, to a 12th-century Indian mathematician named Bhascaracharya who posed the problem in verse in his book *Lilavati*[5]. As we look forward to learning the process of problem-solving, let's imagine a dialogue between a student (**S**) confronted with this problem and a patient instructor (**I**). This isn't quite the sort of problem we're interested in really diving into, but the dialogue serves to illustrate a few points that we'll address in the next chapter.

[5] One translation that can be readily found online is an 1817 translation by H.T. Colebrooke: *Algebra, with Arithmetic and Mensuration*, Brahmegupta and Bhascara, London, John Murray.

1.2.1 *A problem-solving dialogue*

S: Hmm, so lemme see if I understand this. Wait, is a span an actual distance? And a cubit?

I: Yes, they are old-fashioned and inexact measures of length. The cubit at least is about the length of a forearm, and people often take that to be about 10.5 inches. A span is from the tip of the thumb to the tip of your pinky if you stretch your hands out as far as possible, and that's usually interpreted as 9 inches.

S: OK, so a span is 9 inches, so the water lily was 9 inches above the water at first. Then when the wind blows, the bud moves two cubits, so about 21 inches to one side until it is submerged. So we want to know how deep the water is. Umm...

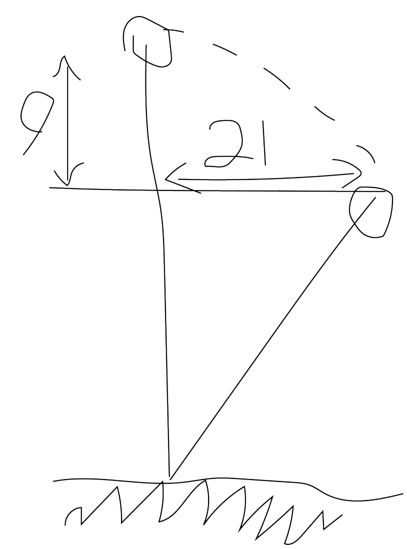
I: Good start.

S: Well how am I supposed to... Hmmm... Alright, so the would the depth just be 30 inches, I mean that's 9 plus 21? No, that doesn't make sense. Shoot... I'm not sure where to start?

I: Do you have a picture in your head of the situation you're thinking about?

S: Yeah, sort of.

I: Do you think you could draw it?



S: Um, I guess so... Yeah... here's the lily bud 9 inches above the water when it's standing straight up... Now when the wind blows it stretches to the side and then the bud is over here, 21 inches away. Wait, is that 21 inches to the right, or 21 inches from here to here [*pointing from the upright bud to the bud at the waterline*]?

I: I think you can assume 21 inches to the right, but that's a really good question, and an important one to get sorted out before you get too far along. Did drawing the picture bring that question to mind?

S: Yeah, definitely. OK, so when the wind blows the lily over there, the stem is slanted like this... and ... so we want the depth of the water, which is this distance [*drawing a vertical line from the lake bottom to the waterline*]... so this makes a triangle, is that what I'm supposed to do?

I: There are no "supposed to's" in this class.

S: Oh, so how will I know when I'm doing it right?

I: There's more than one way to solve the problem, so there isn't just one "right way". If you've interpreted the problem correctly, chosen a solution procedure that doesn't violate the fundamental principles of mathematics, and all your arithmetic is sound, you'll get the correct answer and you'll know it.

S: Oh. But... how do I know if I've got the right answer? You're not going to tell me?

I: Welcome to the real world.

Exercises

1. Make a list of the variables or factors that one might need to take into account when projecting future pheasant populations from historic data such as the Iowa Roadside Pheasant Survey.

2. How do game wildlife managers benefit from making population estimates like the roadside survey?

3. See if you can arrive at a solution for the Lilavati water lily problem. While you're working on this, occasionally step back and think about what sorts of things you are doing to make progress. Are you trying to understand what the problem is about? Trying to appreciate the geometry? Trying to recall algorithms that relate different quantities in a geometric shape? If you feel stuck at any point, what do you think is preventing you from making progress? Record and annotate all of your thoughts and solution attempts, and try not to erase or scribble anything out.

Part I

PROBLEM SOLVING

2

Problem Solving as a Process

2.1 *What is a problem?*

Let's establish this right away: a problem is an intellectual challenge. Solving a problem is then a process of undertaking and overcoming the challenge.

As we have indicated elsewhere, authentic problems are often poorly-structured and vague. They are not carefully-crafted to yield whole-number answers with a few minutes of symbolic manipulation, like those you typically encounter in school. Authentic problems could take hours, days, or even longer to solve, and you may not always know with confidence that you have succeeded because the answers aren't in the back of the book. Problems don't require only the application of a recently-learned method or algorithm; indeed, you may not know at the outset which methods are appropriate for solving the problem. You may not be given all the information needed to achieve a complete solution, or information you are given may be uncertain or incomplete. In short, authentic problems are hard, and that can be frustrating.

Many mathematics teachers and scholars distinguish between **problems** and **exercises**. An exercise prompts you to practice a method that you recently learned or for which you recently studied examples. Exercises usually have **answers** that can be compared with an answer key. A problem is more challenging because there may be fewer cues to the appropriate solution methods and there may be no explicit relationship between these methods and the instruction received.

But wait! Don't close your book (or laptop) and walk away just yet! Having just explained the difficulties, consider the flip side of the problem-solving coin: problems that are authentic are also inherently interesting, particularly when they address contemporary issues or puzzles in your chosen area of study. Solutions and solution methods for such problems are therefore not just an academic dead-end, but can lend themselves to practical applications in the real world. The rewards of achieving a clever and well-justified solution to a practical and interesting problem should outweigh by far the moments of uncertainty, frustration, or disappointment encountered along the way.

Most of the problems discussed in this book are designed to mimic authentic problems, and in many cases are drawn from or inspired by encounters with local researchers and practitioners. As we endeavor to address these problems, we'll often find it helpful to explore simpler problems and exercises to help with sense-making. Thus, our time will be spent moving back and forth from focus problems to auxiliary problems and exercises. Read on through the end of this chapter to understand how this approach can lead to more successful problem-solving.

Problem solving cannot be reduced to a simple recipe or fast and easy method, but in the past 70 years, much has been learned about how successful construction of solutions differs from unsuccessful attempts. A key component is the use of **heuristics**, or habits of mind that are useful in solving problems. The modern idea of heuristics has its origin in the work of Hungarian mathematician George Pólya in the mid 20$^{\text{th}}$ century. Heuristics help guide us in decisions about how to approach a problem. With the help of heuristics and the benefit of experience, we may develop problem-solving **strategies** that lead to successful solutions. We'll begin our study of problem-solving with a brief look at Pólya's method and some of his heuristics and then consider how they might apply to problem-solving in the natural science and natural resource management contexts.

A **strategy** is a definite sequence of steps or operations that leads to a solution.

2.2 *Pólya's method and beyond*

A credentialed mathematician and academic, Pólya was no stranger to the struggles of solving difficult problems. But he was also a teacher and concerned himself with the development of problem-solving skill and intuition in students. He studied his own problem-solving process and that of his professional colleagues and distilled his observations into four essential principles. These principles are general–that is, their use needn't be limited to mathematical problem-solving. The method can be summarized as follows:

Pólya's Method, Condensed and Slightly Revised

1. **Understand the problem**. What is the unknown or target quantity? Is there enough information to find a solution? How is the information that is available relevant to the unknown?

2. **Plan a solution strategy**. How can you proceed from the information available to the unknown? What steps are necessary, and how will the given information be used?

3. **Execute the solution plan**. If the solution plan is chosen well, the implementation of the plan should yield the sought-after result. If unsurmountable difficulty arises, an alternative plan may need to be formulated.

4. **Check the result**. Does the result satisfy the conditions stated in the problem? Is it consistent with expectations or within reasonable bounds? Can you arrive at the same result using a different approach?

These principles may seem obvious, but when the time comes to actually solve a problem it is easy to overlook one or more of them or to lose track of what we are after. Employing this method as a general framework for approaching problem-solving will yield more consistent success and more reliable results.

If it helps to have a mnemonic to remember these, how about UPEC for Understand, Plan, Execute, and Check.

Consider the most common words out of a typical college student's mouth when confronted with a novel problem: "I don't know where to start"[1]. Perhaps the student really means "you haven't yet told me exactly what to do to get the answer". But if the instructor were to point the student toward a solution method every time she was confronted with a challenging problem, she would learn only two things: 1) how to implement algorithms and compute numerical results as instructed; and 2) to relinquish all control of choosing how to approach and solve a problem to somebody else. Sadly, this is often the best outcome of the standard school mathematics curriculum. The worst outcome is that students dismiss math as boring, difficult, or irrelevant. In some cases, a diligent student develops some facility[2] with basic manipulations of mathematical symbols and numbers, but little or no ability to create the frame of mind and methodological structure needed to *begin* and confidently proceed trying solutions.

For the challenge of getting started, Pólya's framework offers **Understand**. What is the problem really asking, and what exactly do you want to end up with as a result? Take, for instance, the pheasant count problem that we introduced in the first chapter: what is

[1] If I had a nickel for everytime I've heard that...

[2] Though the memory of how to use mathematical algorithms certainly degrades with time unless used or reviewed frequently

the unknown in that problem? Essentially, the question we posed there was how many pheasants we should expect to be living in Iowa in the next few years. In some ways, this is a specific reinterpretation of the problem statement, but the simple act of making that reinterpretation not only helps us understand what we are looking for concretely, but our formal statement of it might clarify to colleagues or readers what our solution is driving at.

Since Pólya's framework was developed from the perspective of a mathematician, some of the questions and suggestions pertain mostly to abstract problems. The framework doesn't take advantage of the fact that most of our problems are situated in real-world contexts and involve quantities whose properties can be used as an asset in identifying, constructing, and evaluating solutions. If you're not exactly clear on what I mean by that, go ahead and peak at the next chapter where we discuss the definition and properties of quantities.

On the next page, I have expanded and elaborated upon Pólya's framework and adapted some of the details for problem-solving in natural sciences.

PÓLYA'S FRAMEWORK, and our elaboration of it for problems in the natural sciences, may help us be better organized, but when it comes to applying the quantitative skills we spent years developing in math and statistics classes, we still have little guidance. In Pólya's *How to Solve It*, this is the point where the idea and utility of heuristics was introduced. In the next section, we will introduce and review some generic heuristics that can aid with understanding the problem and inspiring a solution plan.

2.3 A Few Versatile Heuristics

The heuristics we consider here are just a snapshot of the generic methods we might employ in many problems, and some of these are further elaborated in subsequent chapters. These should be some of your most frequent go-to tools for the initial understanding and planning phases, and can be interpreted differently according to the constraints or conditions of each problem. In selecting optimal strategies, the field of options may be narrowed by examining the nature of the unknown. Think about the unknown and what it represents, not only for the focus problem but for any sub-problems or for any goals identified as necessary to solving the focus problem. Could the problem be stated in the form "how much...?", "how many...?", or "is...or not?". If so, the problem might require arithmetic and/or algebraic reasoning. If data is available or supplied and the problem can be stated in the form "what relationship...?" or "how

HEURISTIC: NARROW THE OPTIONS
Use the known properties of the unknown quantity to identify strategies appropriate to the problem.

Solving Ill-Structured Problems

1 **Understand the problem**
Do you understand the problem as it is stated?
Can you restate the problem in your own words?
What is the unknown or desired quantity or output (be specific)? Is it a number? A function? A procedure?
Can you make a drawing or diagram to illustrate how the unknown relates to any known quantities or to the broader problem-space?
Do you already know approximately what the value should be? Can you guess a ballpark or range of reasonable values?
How accurate does your solution need to be? What are the consequences of errors?
What information do you already have?
Is the information that you already have sufficient to solve the problem?
If appropriate, can you write the problem as an algebraic equation with suitable notation?

2 **Plan a solution**. Consider multiple approaches if possible
Have you successfully solved a problem like this before?
Has somebody else documented a solution method to this or a similar problem?
If the problem can be written explicitly as a mathematical statement, do you recognize an algorithm or heuristic that can yield the desired unknown?
If a solution method is apparent, can you assemble all the needed quantities?
If the problem is not immediately solvable, or a solution method not yet apparent...
Can you break the problem into smaller sub-problems that may be easier to solve?
Can you approximate uncertain or unknown quantities?
Could you solve a related auxiliary problem to gain insight?
If data are given, what exploratory analyses could be done to spark ideas?

3 **Execute the plan**
At each step, check to see if the incremental result matches expectations.
Double-check all formulae and algebraic manipulations.
If appropriate, is unit/dimensional homogeneity satisfied?
If you encounter difficulties, revisit the plan and alternatives.

4 **Check the solution**
Is the result reasonable?
Is it consistent with ballpark estimates or benchmarks?
If appropriate, can you substitute the result into the original problem and satisfy the assumptions and conditions?
Double-check algebraic manipulations.
Double-check numerical compuations.
Could you document your full solution with a concise but complete summary of steps, justification of assumptions and methods?
Could a colleague reproduce your approach and find the same solution?

does...change as you vary...?", graphical and statistical reasoning are probably appropriate. If the problem can be stated in the form "how big...?", "what distance...?" or "where...?", then geometric or spatial reasoning could be necessary[3]. There are certainly problems that won't be easily stated in any of these terms, and all options should then remain open. Nevertheless, recognizing common properties in the nature of problems can sometimes narrow down your choices of strategies and make promising solution methods or approaches more apparent.

[3] Each of these types of reasoning and the context-specific strategies that are particularly helpful for them are reviewed in separate chapters in this book.

- **Break the problem into sub-problems.** Complex problems often require multiple steps that can be divided into discrete sub-goals. For example, determining the value of an unknown quantity that is required to solve the focus problem can be considered distinct from solving the focus problem itself. Therefore mapping out the solution in terms of incremental sub-goals can clarify the pathway to a solution.
- **Express conditions algebraically.** Assign symbols to relevant quantities and express the relationship, if known, as an equation relating the quantities. If relationships are not known beforehand, use dimensional analysis to suggest them.
- **Guess the correct answer or solution.** Usually a guess or ballpark estimate is not sufficient if the issue is truly a *problem*, but estimates can still be used to help you recognize if you are on the right track in later computations. If you know approximately what the solution is or what range it should lie within, use this value to check your work. If the solution is not known beforehand but algebraic or practical constraints are available on related quantities, use those quantities to get a ballpark estimate. At this stage, back-of-the-envelope calculations in scientific notation can make quick work of it.
- **Try a few values.** Where some values are unknown but are not the desired quantity, try to solve the problem with a few supposed numerical values. Sometimes the outcome of algebraic relationships is relatively insensitive to the precise value of the quantities included in the relationship. When the relationship is strongly sensitive to unknown or poorly-constrained quantities, identify those quantities as important intermediate goals or sub-problems.
- **Draw a picture or diagram.** When the problem is inherently spatial, such as in the case of habitat or landscape ecology problems, make a map or schematic drawing of the geometric or spatial relationships between known and unknown quantities. To the extent that it is possible, scale distances or spatial dimensions accurately.

- **List all possible cases.** If you're confronted with a logic or simple probabilistic puzzle, make a list or matrix containing possible permutations or combinations.

- **Work backwards.** Where the desired ending condition is known or can be approximated but the steps to reach it are not known, use the end result to help "back out" the steps.

- **Visualize the data.** If data or input values are given and a relationship or summary-statistic is desired, make a graph or diagram from the data. In some cases, visual representation of the data can suggest or substantiate approximate values for the unknowns or can illustrate the form of functional relationships.

- **Solve a simpler problem.** Sometimes a condition or relationship is too complex to solve easily in its full form. In these cases, it may be possible and helpful to simplify the condition to make the problem tractable. You can do this by assuming that an unknown value is known (as in *Try a few values*), by eliminating one or more terms in sums and differences, or by using a simpler statement of the original condition.

The heuristics above are by no means a recipe for success in every situation, but they should be available to you in your repertoire of things to consider. In the chapters that follow, we will elaborate on some of these strategies and add more context-specific tools that can be applied in practical problems.

2.4 *Stepping back*

Before we set you loose on solving problems, it is important to address the mind-set of problem solving. If you have ever uttered the words "I suck at math" or something similar, this section is particularly for you. But as far as I'm concerned, realizing how the mind constructs knowledge and understanding in a problem-solving task is an empowering notion. Alan Schoenfeld, mathematician and math-education specialist, has identified four aspects of the mental process of problem-solving that are essential: Resources, Heuristics, Control, and Belief. Each is necessary and problem-solving cannot or will not proceed without them. First, let's look at what each means, and then we'll consider how they contribute to our problem-solving success.

- Resources

  These are the things you know or understand about the problem domain, constraints on quantities and their representation in the problem domain, and the skills you possess in performing algorithmic procedures.

- Heuristics

  The decision-making tools used to make sense of challenging problems that allow you to make progress or develop insight. Most of the chapters of this booklet are devoted to developing strategies useful in natural science or resource management domains.

- Control

  Control is the management and self-awareness of the problem-solving process. It includes planning, execution and evaluation decisions and and the selection of resources and heuristics for the problem.

- Belief

  Belief includes the set of notions one has about the problem domain as well as one's own abilities or challenges in applying math and statistics to the problem domain; this also includes preconceptions and (mis)understandings that could lead to the use of (in)correct resources and heuristics in a given problem.

Your mathematics and statistics education up to this point has almost certainly stressed resources. Thus, the quantitative resources you bring to a problem consist of all the algorithms and methods you know how to use and your understanding of what they do or mean. Unless you've followed a curriculum through school and college that has deliberately made use of these resources, you've probably forgotten many of them, but re-learning them may not be as challenging as learning them naively. Control and belief are gained from experience, and can build from any foundation of resources and heuristics. The heuristics and strategies themselves can be a bit of a problem though.

You may have been instructed some in the development solution strategies in school, but there's an important distinction between learning what to use and learning when to use it. Some people may argue that it is not a mathematician's prerogative to instruct students in their service courses in more than resources, since heuristics vary from one discipline to another and control and belief grow with experience. This argument is fair, but as non-mathematicians we're left with training in how to implement algorithms, but little idea about how to use those algorithms unless presented with problems that aren't really problems but exercises.

As a result, when the going gets tough,...we get stuck. That's where this course comes in (hopefully to the rescue??). This is your opportunity to work with resources you already have at your disposal, perhaps learn a few more, and to be introduced to strategies for using them in problems that you might encounter in other natural resource courses, in internships, or in your career. By working through these problems in a systematic manner, you'll learn how to control your problem-solving process while building your experience base. I sincerely hope that your belief system evolves in such a way that you become confident that *you can solve quantitative problems too*!

Exercises

1. Think of a challenging problem that you needed help to solve in one of your high school or college courses. It could be a mathematical problem, but doesn't have to be. Describe the problem and reflect on what assistance you needed to arrive at a solution. Were you unable to get started? Did you need help with recognizing and implementing the appropriate algorithms? What was the nature of the assistance that helped you solve the problem? Was it satisfying to arrive at the correct solution?

2. Now reflect on a challenging problem that you were able to solve correctly without assistance. Why were you successful? Were you able to overcome any difficulties or hurdles along the way? Was it more or less satisfying to solve this problem on your own than to solve a problem with help? Why?

3. What resources do you think a person needs to be able to make good predictions of pheasant population over the coming 5 years?

3
Some teaser problems

To illustrate some of the ways that we can employ a deliberate problem solving process, apply general heuristics, and look toward development of solution strategies, I offer some teaser problems. We'll revisit these problems throughout the text where they serve as fodder for more targeted discussions of solution strategies. These problems vary in complexity and are in most cases fairly open-ended. As we've already discussed, these are features of real problems and though we may feel a bit of anxiety about that, we should also realize that this presents *opportunities* for creativity and ingenuity.

3.1 Waterfowl easements

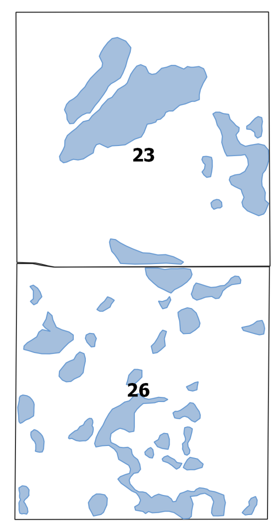


Figure 3.1: Two adjacent sections, each approximately 1 square mile, with different number and size distribution of wetlands.

To have the greatest benefit to migratory waterfowl, is it better secure easements containing many small wetlands or fewer, larger wetlands?

One of the greatest threats to waterfowl populations is loss of habitat. Organizations like Ducks Unlimited and The Nature Conservancy have advocated for preservation and restoration of wetlands in key migration corridors and breeding areas. Among the strategies that organizations like this have used is acquisition of conservation easements, wherein a private land owner agrees to limit development on land containing quality habitat in exchange for benefits like management assistance and tax deductions. In practice, these organizations cannot accept all easement donations but need to prioritize those that will have the greatest long-term benefits to waterfowl. Thus the question: if all else is equal, is it better to prioritize a parcel with many small wetlands, or a parcel of the same size with fewer, large wetlands (Figure 3.1)?

3.2 Herbicide purchase

How much herbicide is needed to kill and suppress regrowth of invasive woody shrubs in a wooded urban park?

A common management problem in urban woodlands is the control of invasive or unwanted understory vegetation. In the midwestern USA, bush honeysuckles (genus *Lonicera*) and European buckthorn (genus *Rhamnus*) are particularly challenging to manage, and the most effective means of control is usually with herbicide. The simplest approach for many of these woody invasives is to cut all the stems and paint stumps with a general herbicide like glyphosate. So what must we know and determine to decide upon the quantity of glyphosate required?

3.3 Deer-automobile collisions

How likely are collisions between automobiles and whitetail deer in your county, and how might they be influenced by changes in deer population management?

If you live in an urban environment without open spaces and wildlife corridors, the answer to this may be easy: zero likelihood. But in many areas both periurban and rural, deer are a familiar sight. In these settings, deer-automobile collisions may unfortunately be all too common. We might infer that the frequency of collisions would be greater where both deer and drivers are more numerous, but can we quantify this?

3.4 Prairie dog plague

Is it possible to prevent the spread of Sylvatic plague through prairie dog colonies?

Yersinia pestis is a flea-borne bacterium that causes Sylvatic plague, an often fatal disease that primarily affects rodents. The black-tailed prairie dog (*Cynomys ludivicianus*), already threatened by land-use change and hunting/poisoning within it's native range in the North American Great Plains, periodically experiences plague epidemics that can decimate colonies. In addition to harming prairie dogs directly, however, plague can be transmitted both directly and through fleas from prairie dogs to their predators, including the endangered black-footed ferret. In isolated populations of prairie dogs, monitoring and active management can potentially be used to reduce the spread of plague, but designing such a strategy requires that we first understand the dynamics of disease transmission within colonies.

3.5 Forest fire losses

How much fuel management is optimal to minimize risk of fire damage in a forest managed for timber production?

In many western forests managed for timber production, historical fire exclusion has led to a buildup of fuel wood, leading to heightened risk of destructive fire. When fires do occur in these settings, high fuel density can lead to destructive crown fires that greatly diminish the value of the timber and can damage or destroy surrounding infrastructure. To combat this, forest managers can undertake fuel reduction treatments such as thinning and prescribed burning, and emergency managers can aggressively fight fires when they do start. However, such suppression and pre-suppression efforts can be costly and can have diminishing returns in terms of cost and resources (Figure 3.2). Is there an optimal amount of fuel reduction and suppression that provides some protection from risk while containing costs?

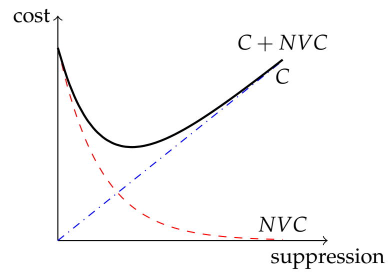


Figure 3.2: Schematic illustration of one conceptual model for the optimal management of fire fuels. Total cost C of management increases with suppression and pre-suppression effort. With little suppression effort however, the risk of high losses or net change in value, NVC, is high, declining with suppression effort. This view of management incentives indicates that overall cost is $C + NVC$.

3.6 Maximized effluent

How much of phosphorus (P) can be discharged from point sources into an urban stream without exceeding total maximum daily loads (TMDLs)?[1]

[1] Inspired by Litwack et al., 2006, *Journal of Environmental Engineering*, 132(4): 538-546.

The U.S. Clean Water Act of 1972 establishes criteria for identifying polluted water bodies and designing plans to reduce pollutant loads and improve water quality. One of the measures available to water resource managers is the establishment of Total Maximum Daily Loads (TMDLs), which typically limit the permissible concentration of a pollutant in an impaired water body. If the goal in this case is to limit the amount of P delivered to downstream water bodies, we need to determine how much P constitutes the maximum allowable concentration.

3.7 Brook trout recruitment

Given data from four consecutive electrofishing passes in an isolated stream reach, what is the stream's age-0 brook trout population?

One method of estimating fish populations in streams is to isolate a stream reach with barriers upstream and downstream of and performing an electrofishing transit of the reach, collecting and measuring each fish retrieved. When extra detail is needed or fish are elusive, multiple passes through the reach may be required, and retrieved fish are removed to a live well or adjacent reach. Through this method, the change in catch between consecutive passes can be used to estimate the actual population within the reach. To isolate the population of age-0 (young of the year) fish, age must be assessed from each fish either directly[2] or indirectly from the length distributions in the catch in each pass.

[2] A nonlethal age estimate can be obtained by collecting and observing fish scales under a microscope

The problems above deal with a variety of practical issues that could be addressed by a natural resource professional. None of them are expressed explicitly as math problems, but quantitative methods could be instrumental in solving each of them. Indeed, this is characteristic of many real science and management problems. Before they become quantitative problems, they must be interpreted and carefully re-framed (Figure 3.3). This is not only an important part of the process, but in many cases is the most challenging and pivotal step. In practice, this must be part of the process of understanding the problem and perhaps even planning a solution strategy. In most cases, not enough information is given up front and the expectations for the kind of solution desired are vague. In this sense, they are ill-structured problems, requiring some digestion and careful re-phrasing before they can be understood as solvable quantitative problems.

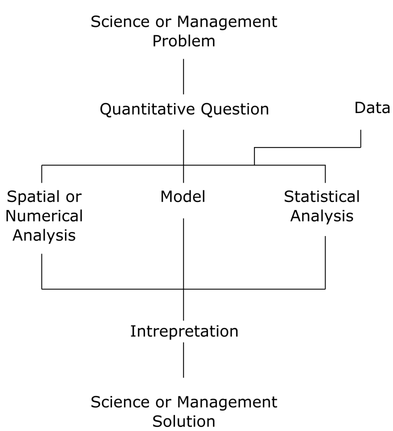


Figure 3.3: A pathway between problems and solutions in practical natural resource management, including various quantitative approaches.

Gathering the information deemed necessary to arrive at the solution is, of course, an essential step in the process. When possible, we should strive to include this step as part of the coursework. However the design of experiments and logistics of making novel field measurements can and should be addressed in their own courses, and we cannot hope to do justice to those concerns here. As a consequence, where the teaser problems are addressed here and in the chapters that follow, we'll either work with hypothetical data or engage real data from the literature or from government documents and web resources.

Importantly, these problems also vary in the degree to which we could ever hope to know that we have the "right" answer. For example, It may indeed be possible to know exactly the maximum P load delivered in Problem 3.6 or the herbicide volume needed in Problem 3.2, provided that we had perfect information on stream discharge and stem basal area, respectively. However, for most of the other problems, confidence that we have a good solution must come from confidence that we have made well-reasoned interpretations and assumptions and used technically-correct manipulations and analysis. Thus, part of the burden of solving such a problem is articulating how the problem is interpreted and how chosen solution strategy follows from that interpretation.

Each of these problems provides hints or implicit cues to what kinds of strategies are appropriate to assemble a solution. For example, there are clear spatial elements to Problem 3.1 and Problem 3.2, but there may also be subtler spatial components in others. By spatial, I mean that we need to incorporate information about how big or how long something is into our solution. Addressing these elements of problems requires some tools from geometry and trigonometry, as well as perhaps the language and conventions of geography or cartography. Therefore, this book contains several chapters exploring the aspects of spatial reasoning that arise frequently in natural resources.

Spatial reasoning is the use of spatial information or relationships, like lengths, areas, volumes, or directions, in the solution of problems.

Not surprisingly, several of these problems point to issues of *how many*, particularly Problem 3.3 and Problem 3.7. In these and other problems, we may wish to characterize individuals or populations in terms of a representative value or distribution of values, or we may need to compare values or proportions. These tasks engage our experience with arithmetic, descriptive statistics, and probability, but in many cases also require that we are careful with units and that we can work efficiently with the extremely large or small numbers sometimes encountered in the sciences. For these issues, we have several chapters devoted to numerical reasoning.

Numerical reasoning, as used in this book, is the manipulation, characterization, comparison, and interpretation of numerical values (such as data) in the service of problem solving.

Aspects of most of our problems can be expressed in terms of the relationship between multiple variables, or relationships between cause and effect. For example, in Problem 3.5 we may need a way to represent the relationship between how much effort is made in fire-suppression activities and their cost. Since we don't necessarily have an idea ahead of time about how much effort is appropriate, an algebraic expression relating the variable C (cost of suppression efforts) to the amount of suppression effort itself (call it S) could stand in and allow us to employ algebraic reasoning as a means to a solution.

Algebraic reasoning is the use of generalized variables and formal relationships between them, rather than numbers, as a means of constraining solutions.

Finally, several of these problems imply that we seek an assessment or prediction of future, hypothetical, or unobservable quantities or events. For example, Problem 3.4 seems to ask whether something is even possible, even though we don't know what that something is. In these cases, finding a useful solution may require modeling. As indicated in Chapter 1, a model is simplified representation of real systems and is constructed for the purpose of exploring cause and effect or functional relationships between variables as system conditions are changed. To model the spread of plague among prairie dogs, then, we may need a way to characterize the transfer of fleas between an infected individual and an uninfected, but susceptible individual. This type of reasoning will often require algebraic constructs and simplifying assumptions, and may therefore require competency in algebraic reasoning. Even so, however, some insights can be gained from model construction even if the detailed relationships between

Modeling in this book refers to the construction of a simplified representation of real or hypothesized systems, often described with one or more equations relating variables to one another and to system properties, and used to explore complex or unobservable phenomena or relationships.

variables aren't formally articulated with equations.

The Parts and Chapters that follow are elaborations and demonstrations of problem-solving strategies employing each of these types of reasoning. Not all will be applicable to every problem, and there isn't necessarily a sequence of interdependence. Therefore, I hope that rather than reading the text in order, that you will consult it as the need for ideas and insight arises in problems you are presented with. I don't provide complete solutions to any of the teaser problems above, but do use parts of them to illustrate concepts and strategies as they arise.

Exercises

1. Which of the teaser problems outlined above do you find most interesting or compelling and why?
2. Using the table on solving ill-structured problems in Chapter 2 as a guide, attempt to UNDERSTAND a teaser problem of your choice from the list above. Write your response to each of the questions under the UNDERSTAND prompt, or write "n/a" if not applicable.
3. Try to write one or part of one of the teaser problems as a more conventional math problem, with only equations and no words. What is challenging about doing this?
4. Write a problem of your own in a format similar to the teaser problems above: a short title, a boldface question, and a short paragraph elaborating on the context or significance of the question. Choose a natural resource topic that interests you.

Part II

NUMERICAL REASONING

4
Quantities in the Real World

In this course, we seek to solve practical problems in natural resource management and ecology, but we focus on the use of quantitative tools in service of this objective. Before diving too deeply into problem-solving, we should ensure that we know what is meant by quantities, quantitative tools, and quantitative reasoning. We will also establish a few conventions for how quantities are represented in science and how quantitative information can be most effectively communicated.

4.1 Quantities in Natural Resources

If we can assess the presence or absence of something, count its number, measure some property that it has, or compare it to another object, it can be *quantified*. That quantified thing is then represented by a **quantity** that is itself now a property of the quantified thing. If that sounds confusing, read on to some of the examples below. A fully-defined quantity has five components:

Properties of quantities

- **Name**: what we call it.
- **Procedural statement**: how it is measured or computed.
- **Number**: numerical value(s) corresponding to magnitude or multitude.
- **Units**: how it is *scaled*.
- **Symbol**: a character that stands for the quantity in equations.

Defining a quantity might seem somewhat pedantic, but it has important implications for what we can and cannot do with it. This contrasts fundamentally with the abstract variables we encountered in high school math. In that setting, there is rarely any reason to question whether it is OK and meaningful to add $3x$ and $8y$, we just do what we're asked. But in the world of real quantities, if x stands for "milligrams of sodium chloride" and y stands for the number of eggs in a Northern Cardinal's nest, it's not so clear that we can perform that addition. Even if we do, it is not so clear what the result means.

In the introductory chapter, we pondered the Iowa DNR's roadside pheasant survey and what it means for pheasant populations across the state. The pheasant count yields a single number each year, for example 23.9 individuals per 30 miles in 2015. We discovered what this quantity means and how it is measured in the Introduction, so we already have most of the ingredients of a fully-defined quantity. We only need a symbol. This is a pretty trivial step in simple problems, where the primary constraint is to make the symbol unambiguous and suggestive of the quantity it represents. Perhaps we should then choose P for our symbol. If we were to prepare a document describing the DNR roadside pheasant survey, once we establish each of the five properties[1] of our quantity, we can thereafter use P with confidence that the information conveyed by that symbol is clearly established: within the context of our document, P would refer to the series of annual estimates of pheasant density according to the method established by the DNR. This formality thereby provides a shorthand name and eliminates any ambiguity in discussion of quantities. Note that this is different than P, which we used previously as the chemical symbol for the element Phosphorus. If we happened to be working on a problem involving both pheasants and Phosphorus, we might select our symbol differently to avoid ambiguity.

[1] Name, procedural statement, number, units, and symbol

As we indicated above, there are many types of quantities that we encounter in science. We can treat presence/absence information as a quantity, measured with a *nominal* scale[2]. Was there a black-capped chickadee on the bird feeder at 3:30 PM? Is there a beetle in the pit-trap we set up in a field? An *ordinal* scaled quantitiy is one in which individual measurements or components are ranked or ordered. In what order did different birds arrive at and depart from the bird feeder[3]?

[2] Nominal-scaled quantities take the form of categories; a few pairs of categories that fit this definition might be present or absent; infected or not infected.

[3] Ordinal-scaled quantities have values according to their rank among the population or data.

In the above cases, distinctions between the scales are clear and obvious, but in others they can be more subtle. Consider the expression of the quantity temperature. We have several scales we can choose from for measuring and expressing temperature. Most important and familiar are the Fahrenheit, Celsius (Centigrade), and Kelvin (absolute) scales. The Celsius scale, for example, is defined according to the freezing and boiling points of water, and the temperatures that those phenomena corresponded to were defined to have values of

0°C and 100°C, respectively. A degree of Celsius is therefore defined on an *interval* scale[4], where a unit of temperature change is 1/100th of the difference between the boiling point and freezing point of water. In contrast, Kelvin is a measure of the absolute thermal energy contained in a substance, in reference to a theoretical state of zero energy[5]. Kelvins are a *ratio* scale[6], with a natural zero point and a unit magnitude that represents the value on a scale relative to that zero point. It can be confusing to distinguish between interval and ratio scaled quantities, but the following example might help illuminate the difference. Imagine that you measured a lake's surface water temperature at one moment in mid March to be 1°C. Now suppose you returned a week later to find that the temperature has increased to 2°C. Wow, the temperature doubled, since 2 is twice as large as 1! Well, no it didn't. Because Celsius is an interval scale, chosen arbitrarily to have a value of zero at the freezing temperature of water, we cannot say it doubled. The issue with this might be even more clear if our first measurement had been -0.1°C (supercooled!) instead of 1°C. Then if we apply the same logic as before to characterizing the change to 2°C, we would have to say that the temperature is -20 times the initial value, which is ridiculous. It's ridiculous because our scale does not have a natural zero point, and instead describes interval differences between a measured temperature and a standardized reference point. In contrast, a temperature of 300 K *is* actually twice as hot as a temperature of 150 K, since this is a ratio scale[7].

[4] Interval-scaled quantities may take negative values.

[5] the Kelvin scale, named for Lord Kelvin (William Thomson), is considered an absolute temperature scale, measuring the absolute thermal energy in a substance.

[6] Ratio-scaled quantities cannot have negative values; a non-positive value indicates an absence of the quantity.

[7] For additional topical perspective on quantities and scaling in ecology, consult a resource like Schneider, D.C., 2009, Quantitative Ecology, 2nd ed., London, UK, Academic Press, 415 p.

It might seem a bit esoteric to define quantities and their unit scales in so much detail, but hopefully you can now see how different scales might entail different rules for what manipulations are or are not meaningful.

4.2 *What Quantitative Reasoning Is*

Using, manipulating, and interpreting quantitative information are the prerogative of professional scientists and natural resource managers. The term **quantitative reasoning** is often used to describe the processes of constructing and making sense of contextualized quantitative information[8]. When we look for patterns in water quality data, interpret pheasant population trends, or estimate the number of person-hours required for removal of invasive shrubs in a forest, we are applying quantitative reasoning. An individual's aptitude for quantitative reasoning is sometimes referred to as **quantitative literacy**.

[8] One researcher describes quantitative reasoning as "sophisticated reasoning with elementary mathematics, rather than elementary reasoning with sophisticated mathematics." (Steen, L., 2004, *Achieving Quantitative Literacy: An Urgent Challenge for Higher Education*, Washington, DC, MAA.)

A guiding premise of this book is that the lifetime value of strong quantitative literacy is greater than that of pure mathematics for natural resource professionals. To be sure, basic mathematical skills are fundamentally essential for quantitative reasoning. But for most professionals outside of engineering and the physical sciences, these skills are mostly learned in secondary school math. What often isn't learned is how to use those skills to make sense of things in the world around us.

4.3 *Units & Dimensions*

Paying attention to units and dimensions is not mere formality. In the scaled quantities we use in science, units and dimensions constrain what mathematical operations are permissible.

Quantities with practical meaning to us will often have units. Consider a few random examples:

- 71 foot tall tree
- 16.4 grams of soil
- 42 snow geese
- 39 breaths per minute
- 385 ppm carbon dioxide

Each of these examples would lose much of their meaning if we only stated the number and not what the number corresponded to: 71, 16.4, 42, 39, 385. We can certainly plug these numbers into equations and work with them in a purely mathematical sense, but they probably no longer have a meaning that we care about. Thus, we need to keep track of *units*. Units can be any standard template or scale according to which we measure a quantity. A foot, for example, though originally loosely defined by the length of a Greek or Roman man's foot, is now defined with reference to a meter, which in turn is defined according to the distance light travels in a vacuum during a pre-defined time period. Likewise, grams and minutes are units of mass and time defined according to internationally standardized references.

Dimensions are sometimes conflated with units, but the term is distinct and more broad. In casual usage, the term "dimensions" often implies that we wish to know how large something is - in other words, its length, area or volume. In the physical sciences, dimensions are a way to group different types of units that can be simply scaled with one another. Inches and centimeters are both lengths. No matter what object you measure – perhaps a young brook trout's total length – if it is an inch long, it is by simple unit conversion also 2.54 centimeters long. Both are units of length. But just because I find that the fish is 1 inch or 2.54 centimeters long, that does not necessarily tell me how heavy it is. So mass is a different fundamental dimension than length, and whether I measure mass in grams or slugs, it has dimensions of mass. In fact, most of the quantities that scientists deal with are composites of only three fundamental dimensions: **mass**, **length**, and **time**. Sometimes we use the symbols $[M]$, $[L]$, and $[T]$ in square brackets to denote when we are talking about dimensions. We can also have dimensionless quantities which we can (for reasons that we'll see below) denote with [1], and which we'll often encounter when we are dealing with nominal, ordinal, or multitude scales.

Most quantities in the natural sciences have dimensions involving mass $[M]$, length $[L]$ and time $[T]$

A quantity made up of any combination of the fundamental dimensions can be called a *derived quantity*. Speed is an example of a derived quantity because it implicitly contains two quantities, a distance or length traveled during a period of time, $[L\ T^{-1}]$. Note that we haven't said whether we are measuring length in meters or feet or whatever, nor have we said that time is in seconds or hours or millennia. Dimensions are more like categories of units, and we can convert from one set of units to another within each category by performing multiplication or division. Energy is also a derived quantity, and can be expressed in units of Joules or Calories. But regardless of the units employed, any quantity of energy has dimensions of $[ML^2T^{-2}]$. We just need to be careful if we are working with equations that we don't mix units, or we'll be left with gobbledygook.

4.4 *Comparing apples and oranges*

We all learn in junior high or high school science to pay attention to units, and for good reason. Keeping track of units and making sure that our units are consistent in any computation we do with physical quantities can prevent costly mistakes. I'm not terribly pleased about it, but I probably spend a few hours a month painstakingly performing unit conversions in my research – indeed it may be the most frequent type of computation I do, but I do it because I know it is important.

Let's consider a simple contrived example, similar to one that we will encounter later this semester: suppose you are told that the soil in your back yard has 15.84 ounces of lead per metric ton of soil. Should you worry? Well, according to EPA guidelines, remediation is recommended if lead concentrations in residential yards exceed 400 parts per million. Parts per million is a dimensionless unit similar to a percent but much smaller and it can correspond to a mass (or volume) of one substance contained in a mass (or volume) of some mixture of substances. Our measurements are already both masses, so 15.84 ounces/1 ton is $[MM^{-1}]$, and is already dimensionless, but our units are not consistent. We could fix this in several different ways, but one simple one might be to convert tons of soil to ounces, so that we have ounces over ounces, yielding consistent units. We could spend lots of time (not wisely, I would argue) setting up fractions on paper to figure this out (how many ounces in a pound, how many pounds in a kilogram, how many kilograms in a ton. . .), but as long as we have access to a computer we might as well use it. Plenty of apps and websites will do basic unit conversions for you, including google itself. Employing one of these, we see that there are 35274 ounces in a metric ton, and now we can express our measured concentration as 15.84 ounces / 35274 ounces, or about 0.00045. To convert from this decimal value to parts per million, we just need to multiply by one million or 10^6 (much like you would multiply by 100 to get a percent), giving us roughly 450 ppm lead. So at 450 ppm, the lead concentration in our soil exceeds the threshold that the EPA has deemed safe, but is not cause for great alarm.

4.5 Problem solving without numbers

Dimensions are so useful in science and engineering that there are entire subdisciplines devoted to dimensional analysis. While these folks are usually using dimensional analysis as a technique for extracting relationships between variables in complex equations that cannot be solved directly, we can also use dimensions to help solve simpler problems and catch errors or inconsistencies. This is because physically-meaningful equations that describe relationships between quantities must be **dimensionally homogeneous**[9]. That is, the dimensions on one side of the equation must be the same as those on the other side. If we know the units of every quantity in a mathematical relationship and can figure out the corresponding dimensions, we can compare dimensions on either side of the equation to verify that it is dimensionally homogeneous. This can feel alot like solving an algebraic equation without actually using any numbers, just combinations of $[M]$, $[L]$ and $[T]$.Doing this *after* making some algebraic manipulations on an equation is a handy way of checking for mistakes! Similarly, if we are uncertain of the dimensions of a constant or variable in an equation, we can solve for that constant on one side of the equation and the dimensional grouping on the other side, through the requirement of dimensional homogeneity, will apply to the unknown.

[9] Dimensional homogeneity requires equations, if expressed only in terms of the units representing each quantity, to remain equal

HEURISTIC: DIMENSIONAL HOMOGENEITY
In constructing or manipulating algebraic relationships, enforcing and verifying dimensional homogeneity can yield insights and catch errors.

Rules for algebraic manipulation of dimensions are straightforward. Multiplying, dividing, or raising to a power any quantities of any dimensions is permissible, and the resulting quantity has dimensions that are the appropriate product, quotient, or power of the original dimensions. Adding and subtracting may only be done between quantities with identical dimensions (and later you should verify that the units are also identical). Let's consider an example where we wish to find the dimensions of the entities in a simple equation describing the accumulation over time of insects (measured in mass) in a pit trap:

$$m = m_0 + kt \tag{4.1}$$

where we know m is mass and so has dimensions $[M]$ and t is time and so has dimensions $[T]$. What are the dimensions of the other two variables? Well, assuming the equation is dimensionally homogeneous and knowing that the left-hand side has dimensions $[M]$, the right-hand side must work out to have dimensions of $[M]$ as well. Furthermore, since added quantities must have identical dimensions, we know that each of the two terms added together on the right hand side must have dimensions of $[M]$. So m_0 is a mass, and k must be something that, when multiplied by a time with dimensions $[T]$, gives a mass. Therefore, k must have dimensions of $[MT^{-1}]$, or mass per time.

Rearranging equations algebraically with only dimensions or units rather than numbers is a useful way to find the units or dimensions of unknown variables.

4.6 Communicating Quantitative Information

We all construct our understanding of quantitative information a little bit differently. I personally grasp quantitative patterns or relationships best if I see it in a graph. Others might have an easier time reading or listening to a verbal description of patterns and relationships. Still others will be more moved by equations or lists of numbers. We can understand and communicate quantitative information in all of these ways – and maybe more (we can encode quantitative information in sound, right?)! But to be well-rounded scientists and managers, we must be able to create and interpret each form. The diagram in Figure 4.1 illustrates these four ways of communicating quantitative information.

Here's what each of the nodes of the triangle mean:

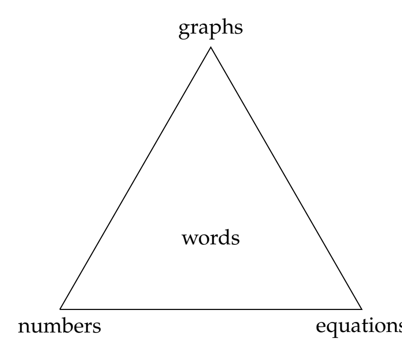


Figure 4.1: Ways of communicating and understanding quantitative information.

- **Graphs, images, figures**: any visual display of quantities or relationships between them. Single-variable statistical charts (e.g., histograms) and two-dimensional graphs (i.e., x-y scatter-plots) will be the most frequently encountered, but maps are another example. Often the most efficient way to demonstrate patterns or large volumes of data.

- **Numbers, in lists or tables**: for numerical information, the most direct, precise, and unambiguous way of communicating quantities that aren't too numerous (i.e., a short list).

- **Equations, inequalities, or proportionalities**: a formal and precise way to state hypothesized, derived, or observed relationships between quantities.

- **Words, concepts**: descriptions and interpretations, either standing alone or to accompany another form of expression.

In technical reports, it is good practice to employ at least two or three forms of quantitative expression, where one form will always be words. Words are in the center of the triangle because they must be used to link the other forms conceptually, and without them we cannot claim to be understanding and communicating effectively. Furthermore, as professionals, we cannot simply provide charts or tables of data and expect them to speak for themselves. Part of the role of scientists and managers is to interpret quantitative information and make decisions or recommendations based on our interpretations.

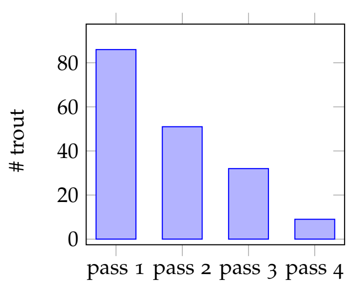


Figure 4.2: Bar chart showing the number of brook trout removed from a stream reach during each of four consecutive electrofishing passes.

pass #	trout removed	effort
pass 1	86	35 min
pass 2	51	29 min
pass 3	32	28 min
pass 4	9	31 min

Table 4.1: Brook trout removal from four electrofishing passes in one stream reach, including information about the number of minutes (effort) elapsed during each pass.

4.6.1 *Example: brook trout recruitment (teaser Problem 3.7)*

The data from electrofishing studies of fish population and age-structure can be quite simple. A typical set-up would begin with blocking the channel upstream and downstream with fences or nets to prevent fish from entering or leaving the study reach. Then fisheries technicians would slowly traverse the study reach with one person applying the electrodes through the water and a second collecting the stunned fish into a bucket or live well. Captured fish are then measured, weighed, aged (if desired) and returned to the water. In depletion surveys, the fish are returned upstream or downstream of the blocking nets, so as to avoid immediate re-capture. Subsequent passes would operate similarly, and the presumption is that with each removal, the number of fish remaining in the study reach is diminished. Figure 4.2 and Table 4.1 both show data from four consecutive passes through a small trout stream.

The bar graph (Figure 4.2) provides a very simple visual indication of the change in the number of fish captured in each pass. Graphs like this can be extremely valuable for efficiently conveying trends or relationships between quantities. However, they often don't allow readers or viewers to know precisely what the values of the shown quantities are. Furthermore, graphs with too many different variables can become overly complex. If we wish to communicate precise values of quantities, particularly where there are multiple different kinds of variables, data tables like Table 4.1 are perhaps the best option.

Given only the information presented already about the brook trout electrofishing catch, communicating through equations is probably unwarranted. However, a textual narrative is essential for communicating the nature and meaning of the data presented in this example. Suppose you were unfamiliar with electrofishing and depletion methods for fish population assessment, would the bars in Figure 4.2 and the numbers in Table 4.1 tell a clear story? The figure captions and the paragraphs above are essential for drawing meaning from the figure and table. This is why words are at the center of the triangle in Figure 4.1. And this is also why it quantitative problem solving is a writing-intensive endeavor. If we are unable to communicate clearly about our ideas, strategies, results, and conclusions, most of the effort is for naught.

Exercises

1. From our bulleted list of examples at the beginning of section 4.3, what are the units and dimensions for each quantity? Write them with square-bracket notation.

2. In population studies, it is recognized that the number of individuals captured depends strongly upon how much time (effort) is spent in active pursuit. As a result, a better variable to quote than the number of trout caught is the number of trout per unit effort, or catch-per-unit-effort (CPUE). If one minute is defined as the basic unit of effort for the data in Table 4.1, convert the data to a new variable, catch (trout removed) per unit effort (minute). Plot the result in a bar graph similar to Figure 4.2 and create a table with this additional variable alongside the other three.

3. What are the *dimensions* of the new variable created in #2?

4. What kind of quantity is shown on the horizontal axis of Figure 4.2? How do you think this constrains appropriate ways to visualize these data graphically?

5
Working with Numbers

Among the most fundamental operations we do with quantities is arithmetic. We can encounter the need for arithmetic in any phase of problem solving, from making a ballpark estimate in the UNDERSTAND phase to computing and double-checking a final result in the EXECUTE and CHECK phases. Once we have a solid grasp of the operations that are allowable and those that aren't – for example, is it OK to add or subtract quantities expressed in different units or on different scales? – we may get down to business with performing basic operations.

Arithmetic according to Wikipedia: a branch of mathematics that consists of the study of numbers, especially the properties of the traditional operations on them – addition, subtraction, multiplication and division.

Most of us probably feel comfortable with most of these operations, at least when they concern simple numbers. However, it becomes easy to make errors or overlook important steps when we're dealing with extremely large or small numbers, or when unit conversions become necessary. One setting in which we often encounter such difficulties is in working with proportions, including concentrations, ratios, and percentages. Though quantities like these are often conceptually simple, working with them and converting among ways of expressing them can be challenging. This chapter highlights some concepts and techniques for working with these sorts of unwieldy numbers so that we can work confidently, avoid simple mistakes, and even catch more complex ones.

We begin with a method for doing arithmetic that can be used to simplify computations, or to approximate solutions when a back-of-the-envelope computation is all you need. The method is particularly powerful when computations involve very large or very small numbers. As such, it can be useful for making ballpark estimates in the early stages of problem-solving. Our method makes strategic use of scientific notation, which you've probably encountered in secondary science classes. The philosophical basis of scientific notation also leads to the notion of **order of magnitude**, a concept that can be useful for comparing quantities as well as for judging the the appropriateness of estimates. Along the way, we'll compare some ways of expressing normalized quantities like concentrations and proportions, and review the rules for arithmetic with exponents.

5.1 Scientific Notation

In high school chemistry, we learn that there are more than 602 sextillion molecules in a mole of a chemical substance[1]. But we don't normally see Avogadro's constant written as some number of sextillions, nor do we see it elaborated with all of the 24 digits necessary to write it in integer form: it is difficult to keep track of all those digits when writing them, and even more difficult to keep track when you're reading or comparing different numbers. Instead of writing the entire number out, we use the shorthand of scientific notation, where Avogadro's constant looks more like 6.022×10^{23}. In general, scientific notation has the form:

$$a = 10^b$$

where a and b are sometimes called the **mantissa** and **power**, respectively. So Avogadro's constant has a mantissa of about 6.022 and a power of 23, which is equivalent to saying that the complete quantity has 23 digits after the mantissa[2]. Obviously this is a very large number. We can just as easily express very small numbers with scientific notation. An *e. coli* bacterium is roughly 2 μm (micrometers) long, which is 2×10^{-6}m. Here, the power of -6 indicates not that it's a negative number (it would be absurd to say something has a negative length, because length is a ratio scale!), but that it is smaller than 1 and that there should be 6 digits to the right of a decimal point if we wished to express it as a decimal number. So we could express this equivalently in a few ways:

$$2\mu\text{m} = 0.000002\text{m} = 2 \times 10^{-6}\text{m}.$$

Note that these equalities both amount to unit conversions, but the second equality is specifically a conversion to scientific notation.

Negative exponents indicate numbers smaller than 1, and there are occasions where it can be helpful at times to can write these as fractions. When we have a quantity expressed in scientific notation with a negative exponent 10^{-b}, that is equivalent to the same quantity *divided by* 10^b. Therefore, another way to express the length of *e. coli* is:

$$2 \times 10^{-6}\text{m} = 2 \times \frac{1}{10^6}\text{m}.$$

So dividing by 10^6 is the same as multiplying by 10^{-6}. Notice here that the **order of operations** is important. By convention, exponents

[1] 602 sextillion, or 6.022×10^{23} is Avogadro's constant, the number of molecules in one mole of a chemical substance

[2] A related issue is that of **significant digits**. Scientific notation allows us to clearly specify how precise we are claiming to be through the number of digits included in the mantissa: in this case, 4.

In the standard order of operations, parentheses take precedence, then exponents, then multiplication or division, and finally addition and subtraction.

take precedence over multiplication, division, addition and subtraction. So we don't divide by the mantissa (2) when we express this quantity in fractional terms because the only thing that is raised to the exponent is the base, in this case 10. We *could*, however, move the mantissa to the denominator with its 10^6 by taking its reciprocal, right? That's another way of invoking the old grade-school rule: dividing by a number is the same as multiplying by it's reciprocal. In this case, we'd end up with an equivalent value for the length of *e. coli* that looks like

$$2 \times 10^{-6}\mathrm{m} = \frac{1}{0.5 \times 10^{6}}\mathrm{m} = \frac{1}{5.0 \times 10^{5}}\mathrm{m}.$$

Notice that in the last step we've borrowed a "ten" from the power to make the mantissa greater than 1: this is by convention[3]. A general rule for expressing a quantity in scientific notation is to have one nonzero digit before the decimal point in the mantissa, and as many significant figures as appropriate for the problem to the right of the decimal. So we could express the *e. coli* length as 0.2×10^{-5} m or 200×10^{-8} m, but in most cases that is bad form. We shall see below, however, there are times when doing arithmetic by hand can be simplified by temporarily expressing quantities in such an unconventional way.

[3] Quantities expressed in scientific notation should have one nonzero digit to the left of the decimal point.

A useful concept in working with really large or really small numbers is the **order of magnitude** of a quantity. In obtaining a ballpark estimate of a quantity or in computing something using only very rough approximations for the input values, it may be unnecessary or inappropriate to worry about being off by a factor of 2 or so. We might be satisfied knowing that the result is "a few thousand" or "a coupe hundredths". If we're using scientific notation, this is equivalent to ignoring the mantissa and just citing the base and power. So instead of saying that an *e. coli* is 2×10^{-6}m, we can say it is *on the order of* 10^{-6} m long. This kind of reasoning is particularly useful in comparing multiple quantities. A grain of coarse sand, for example, is on the order of 10^{-3}m in diameter, so it is three orders of magnitude larger (-3 is three more than -6) than an *e. coli* bacterium. Once we wrap our minds around what that means (three orders of magnitude is a factor of 10^3, or a thousand!), comparisons can be enlightening in assigning quantitative "importance" to different variables in an equation.

The order of magnitude of a quantity is essentially the value of the exponent when expressed in scientific notation.

THE FACT IS, IN NORMAL COMMUNICATION about the length of *e. coli*, we'd probably stick with 2 μm as a clear way to express it in written text. Most of the alternative ways above are more clumsy in writing, and certainly the last few equivalent expressions above are not intuitive (we only went there to demonstrate the technique!).

However, in comparisons with other qantities or when performing computations with other quantities that are expressed in different units, it is usually smart to convert all quantities to a uniform system of units, like the systeme internationale, or SI.

5.2 *Normalized quantities*

In the sciences, normalization of quantities often refers to the process of dividing some scaled quantity by a standard, total, or reference value of the same quantity. Consider some schemes form normalization that you are already very familiar with. A **percentage** is a normalized quantity, determined by dividing some number that represents a subset of a larger collection by the total number in the collection and then multiplying by 100%. For example, suppose we have tested 360 white-tail deer carcasses (from road-kill and hunter harvest) for chronic wasting disease (CWD) and find that 83 are positive. Given this data, we can all agree that the percent of the sampled population infected with CWD is:

$$\frac{83}{360} \times 100\% = 23.0556\% \tag{5.1}$$

En route to computing this, we created the **ratio** 83 to 360, which is around 0.23 if you simplify it with your calculator. As with many such ratios, we can choose from a variety of different but equivalent ways of expressing this quantity. We could just express it as the ratio of two whole numbers like 83:360[4], or as the fraction:

$$\frac{83}{360}. \tag{5.2}$$

Or as we've already seen, it is simple to write it as a decimal number (0.230556). But since we encounter percentages frequently, we may more readily appreciate it expressed as a percentage. For the present purposes, we could describe a percentage as "parts per hundred", since it is just the same ratio scaled to an arbitrary reference value of 100. In other words, for every hundred deer in the sample, about 23 have CWD. Expressing a quantity in "per mil" is closely analogous, except instead of multiplying by the factor 100% we'd multiply by 1000‰(that's the per mil symbol)[5]. In this case, we'd end up saying that about 83/360 ×1000‰= 231‰(or 231 per thousand deer) are infected. To make this even more absurd, we could express the same information just as easily as parts per million (ppm) or parts per billion (ppb) following a similar tactic. Each of these ways of expressing a normalized quantity is arithmetically equivalent, but implies a different realm of precision about the quantity of interest and the scope

[4] this is the way we usually express a map scale, like 1:24,000. See Part III of this book for more on that issue.

[5] Although not common in many disciplines, isotope concentrations are often expressed in ‰, where the reference value is the isotopic ratio of a standard substance.

of its possible values. We'd likely never talk about deer in parts per million, but we might talk about lead concentrations that way!

Other types of normalized quantities in science include frequencies, concentrations, and probabilities, to name a few. The quantities may be expressed somewhat differently, but in most cases there is a comparison being made between values of the same dimensions (and often the same units!). Indeed this is sometimes a simplifying strategy: when you normalize a quantity to a standard of the same units, details about the specific units by which the quantities were measured can be discarded. Often this is a good thing. For example, when we use a map that is scaled at, for example, 1:24,000, we are not told what units that ratio was constructed with, because it doesn't matter! If you use a ruler to find that the map distance between two features on the map is 2 inches, that distance in the real world is equal to $2 \times 24,000 = 48,000$ in. It doesn't matter whether your ruler is ruled in inches, centimeters, furlongs or rods, the quantity you measure on the map only needs to be multiplied by the scale factor (24,000) to find the true distance! As we've seen, however, neglecting the specific units used to derive a normalized quantity can also be the cause of some confusion (is the concentration of one substance mixed with another computed on the basis of their masses, volumes, or something else?). It becomes a good thing if the *procedural statement* for the quantity is either made clear or is known by convention.

How do we use a normalized quantity to our advantage? Suppose I **extrapolate** from our sample of CWD in deer carcasses to predict that 23% of the deer in the entire county are infected with CWD. If we take for granted that my science is good, all we need to know to find out the number of CWD-infected deer in the county is the total number of deer in the county, N_{deer}. Once we recognize that the ratio of infected deer to total deer is 0.23 (23% of the total population of 100%), we need only perform a simple multiplication:

$$\frac{N_{CWD}}{N_{deer}} = \frac{23\%}{100\%} = 0.23 \tag{5.3}$$

$$N_{CWD} = 0.23 N_{deer} \tag{5.4}$$

Thus, the benefit of expressing the number of infected deer as a normalized quantity (assuming our 23% assertion is accurate) is its generality. We can write a simple relationship like Equation 5.4 and, as long as the relationship remains valid, apply it on any relevant scale[6].

[6] Recognizing how far one can safely scale up from a representative sample is a rich, but complex issue.

The process of re-scaling a ratio (or other normalized quantity) is sometimes called **proportional reasoning**, and is one of the key strategic processes in probability, and as we'll see in the next chapter, it is the foundation of much of trigonometry. The construction of abstract triangles in the service of problem-solving is usually a means of comparing the ratios of two lengths or distances.

There are some oddball normalized quantities in science that are frequently expressed in inhomogeneous units, either as a consequence of their very high or low intrinsic magnitudes or due to the procedure used to measure them. One example is slope in the context of river channels or footpaths, which are often less than 1%. Because typical channel slopes are so small, it is common to see slopes expressed in units of "feet per mile" or "meters per kilometer". They are still normalized quantities, but the inhomogeneous units must be stated explicitly. Similarly, concentrations of solutes or suspensions are sometimes expressed in units like mg/L (milligrams per liter), where the dimensions are a weight per volume. This is convenient because of the relative simplicity of weighing a solid component added to (or isolated from) a volume of liquid. On the other hand, concentrations of substances like dilute hydrochloric acid (HCl; often used in soil chemistry) are are often described as percentages: 5% HCl usually means a mixture in which 5% of the total volume is pure HCl and the remaining (100-5)% = 95% is pure water. Again, this makes sense because when mixed, both components are liquids and their volumes are simple to measure.

5.2.1 Example: maximized effluent, (Problem 3.6)

Phosphorus (P) is a limiting nutrient in many freshwater ecosystems[7]. That means that primary productivity is limited by the availability of P, and that excessive loads of P from fertilizer runoff or municipal and industrial wastes can promote excessive productivity and eutrophication. Thus, we are often seeking ways to reduce the inputs of P into surface waters.

[7] A great review of nutrients in terrestrial ecosystems can be found in Weather, K.C., D.L. Strayer, and G.E. Likens, 2013. *Fundamentals of Ecosystem Science*, Academic Press, Elsevier Inc.

P concentrations in water are often expressed in mg/l, so they are among those normalized quantities that are not dimensionless. A given concentration in mg/l can be visualized as the mass of solute that could be hypothetically extracted from a volume of water, if we somehow had a perfect P-filter. No such filter exists, so not only do we need a different way of measuring P[8], we need more clever ways to extract P from water if it does get in there.

[8] In practice, measuring dissolved P is most efficiently done using a "colorimetric" method wherein a reagent is introduced to a dilute P solution, resulting in the development of a blue color in proportion to the P concentration.

The TMDL selected for P in surface water bodies depends on the designated uses (drinking water? swimming?) of the water bodies in question, but are often on the order of 0.1 mg/l. It's worth remembering that this means that *for every one liter of water*, we should have no more than 0.1 mg of P. So if we happen to take a two-liter sample of water in a water body under this TMDL, we should find no more than 0.2 mg P in that sample, as that (0.2 mg divided by 2l) corresponds to a concentration of 0.1 mg/l.

5.3 *Tricks with scientific notation*

As we've already discussed, simple order-of-magnitude computations can be very informative, particularly in the early phases of problem-solving. This is an occasion when scientific notation can really be useful! To deftly manipulate expressions with scientific notation, it is helpful to remember some key rules for working with exponents.

HEURISTIC
Get a ballpark or order-of-magnitude estimate by hand using scientific notation

RULES FOR MANIPULATING EXPONENTS

$$x^0 = 1 \qquad x^1 = x \qquad x^{-1} = \frac{1}{x}$$

$$x^a \times x = x^{(a+1)} \qquad \frac{x^a}{x} = x^{(a-1)}$$

$$x^a x^b = x^{(a+b)} \qquad \frac{x^a}{x^b} = x^{(a-b)}$$

$$x^{-a} = \frac{1}{x^a} \qquad x^a = \frac{1}{x^{-a}}$$

$$(x^a)^b = x^{(a \times b)}$$

When confronted with problems where multiplication or division of very large or very small numbers might be involved, we can set the problem up in scientific notation to make things simpler. Consider the simple example of determining how many milliliters (ml) are in a cubic meter of water. One thing that is useful to know is that a ml is the equivalent of a cubic centimeter (cm^3). And we also know that there are 100 ($= 10^2$) cm in a linear meter (m). So how do we determine the number of cm^3 in a m^3? Recall from earlier that if we are converting between, for example, one set of squared units to another set of squared units, we need to square the conversion factor for the linear units too! So for this problem, since there are 10^2 cm in every m:

NUMERICAL BENCHMARKS: VOLUME
1 ml = 1 cm^3
1 m^3 = 1000 l

$$1 \text{ m}^3 = (10^2)^3 \text{ cm}^3 \tag{5.5}$$

Using one of the above rules for exponents to modify the right-hand side of this relationship, we can find that:

$$1 \text{ m}^3 = 10^{(2\times3)} \text{ cm}^3 \tag{5.6}$$

$$1 \text{ m}^3 = 10^6 \text{ cm}^3 \tag{5.7}$$

and we have our result. There are 10^6 cm^3, and therefore 10^6 ml in a cubic meter! It would be just as easy to look up the conversion online, but the same basic approach can be readily applied to more complex problems with murkier solutions. In the next section we'll consider a more challenging and engaging example that can be worked out with a similar strategy.

5.3.1 *Example: Mercury in fish*

The RAFT program[9] (Regional Ambient Fish Tissue) is an EPA effort to monitor concentrations of several harmful toxic substances in fish in the state of Iowa. This problem concerns the (slightly idealized and modified) values of mercury (chemical symbol Hg) detected in smallmouth bass sampled from two locations in Iowa. Samples of fish tissue were obtained as "plugs", taken from live fish in a manner similar to a biopsy. Typical plug samples weigh 50 mg. The criteria for issuing fish consumption advisories are shown in the table below. Plugs from smallmouth bass in Lake Wapello, IA contained on average 0.06 μg of Hg, while plugs from smallies in the Maquoketa River contained 0.01 μg Hg. Should there be consumption advisories for either waterbody?

[9] Find more information about the EPA RAFT program by searching EPA raft on the web.

Hg concentration	**advisory**
< 0.3 ppm	no restrictions
> 0.3 to < 1.0 ppm	1 meal/week
≥ 1.0 ppm	do not eat

A simple solution method

A useful first step is to identify the desired result. We'd like to find a Hg concentration in each fish in the same units that the advisory guidelines use: parts per million or ppm. This is a normalized and dimensionless, derived quantity. A second helpful step is therefore to express the key data in uniform units so that we can normalize them in dimensionless form. Our Hg measurements are in μg, which is 10^{-6} g, while our plug mass is in mg, which is 10^{-3} g. It doesn't really matter whether we convert everything to grams or something else, but grams is straightforward. So now me construct the ratio that expresses how much mercury there is, by mass, in our fish tissue sample (using Lake Wapello values as an example):

HEURISTIC:
Convert to uniform system of units

$$\frac{0.06 \times 10^{-6}}{50 \times 10^{-3}} \frac{\text{g}}{\text{g}} \tag{5.8}$$

Simplify this by cancelling units and expressing each quantity in proper scientific notation:

$$\frac{6.0 \times 10^{-8}}{5.0 \times 10^{-2}} \tag{5.9}$$

Using rules for division in exponents with a common base, we can simplify this:

$$\frac{6.0}{5.0} \times 10^{(-8)-(-2)} \tag{5.10}$$

The exponent therefore becomes -6, which you recall is the base for a "parts per million" ratio. We can simplify the fraction $6/5$ either directly on a calculator, in our heads[10], or by multiplying both numerator and denominator by two ($= 12/10$) and dividing by 10 to get 1.2:

$$6/5 \times 10^{-6} = 1.2 \text{ ppm} \tag{5.11}$$

So the result for Wapello is 1.2 ppm, which exceeds safe limits for consumption. For the Maquoketa River, the Hg concentraion is only 0.2, so it is safe to eat and no advisory need be issued.

[10] One great benefit of using scientific notation is that computations can be approximated easily by hand!

5.3.2 *Example: forest fire losses (Problem 3.5)*

Let's use some of the above techniques and strategies to make some ballpark estimates about the value of timber that could potentially be lost in a forest fire, following the teaser problem in Section 3.5. This could give us at least a starting point for imagining where the curve NVC starts from on the left-hand side of Figure 3.2. Since no specific information is given about the size of the property, we need to make and explicitly state an assumption. Let's suppose for now that the property has an area of 1000 hectares, since that number is both reasonable for a single-ownership land parcel (this would be a bit less than 4 square miles) and is easily scaled. Let's also assume that this forest in the absence of any fuel reduction effort is overstocked, with perhaps 30 $\text{m}^2\ \text{ha}^{-1}$ of basal area[11]. Using timber cruising charts, this basal area would yield about 30,000 board feet per hectare[12].

HEURISTIC: Not enough information given? Make and state explicitly a reasonable and potentially-scalable assumption. If appropriate, choose values that can easily be scaled, like 1 or 10.

[11] basal area, usually given in $\text{ft}^2\ \text{ac}^{-1}$ (square feet per acre) or $\text{m}^2\ \text{ha}^{-1}$ (square meters per hectare), provides a quick glimpse of the amount of standing timber on an area of land.

[12] One board-foot is equal to about 0.00236 m^3 of wood.

To get a ballpark estimate of the value of this timber then, we need to find the going price per board-foot of our timber and then scale this up with the timber volume and property area. A reasonable guess for the price for softwood saw-logs would be 0.20 US dollars (USD) per board foot[13]. So our computation becomes

[13] A web search for "sawlog prices" can give you some idea of how this varies by place and time.

$$NVC(0) = 1000 \text{ ha} \times 30000 \text{ BF/ha} \times 0.20 \text{ USD/BF}.$$

We can do this computation relatively quickly in a calculator, but there is a risk of typing in the wrong number of zeros and making an important error. However, if we convert these quantities to scientific notation and rewrite the equation we can do the math in our heads. The parcel area is 1×10^3 hectares, the wood volume is 3×10^4 board feet per hectare, and the value is 2×10^{-1} USD per board foot. So we may re-write the computation as

$$NVC(0) = 1 \times 10^3 \text{ ha} \times 3 \times 10^4 \text{ BF/ha} \times 2 \times 10^{-1} \text{ USD/BF}.$$

Since all these quantities are multiplied together, we can rearrange (by the commutative principle for multiplication) and group the mantissas together, put the powers together, and put the units together.

$$NVC(0) = 1 \times 3 \times 2 \times 10^3 \times 10^4 \times 10^{-1} \text{ ha BF/ha USD/BF.}$$

Multiplying the mantissas through we get 6, and using the rules for manipulating exponents (see the next section!) the exponents are added together $(3 + 4 + -1) = 6$. Canceling units, we see that USD is the only remaining unit. So our ballpark solution is that the value of the standing timber in this 1000 ha parcel is 6×10^6 USD, or about $6 million.

Exercises

1. The Hg concentrations measured in the RAFT program problem were taken from "keeper" size smallmouth bass, roughly 35 cm long. A few scattered measurements from larger and smaller bass indicated that there was some systematic relationship between Hg concentrations and fish size at each site, but not across sites. What systematic relationships would you *predict* to be present in fish of different sizes? What quantities might be relevant to this problem? Formulate a testable hypothesis for expected systematic variation in smallmouth bass tissue Hg concentration as a function of fish size.

2. [14] Leucism is partial albinism, manifested in penguins as a lack of (or substantial reduction in) pigment in plumage. A study of three species of penguin (Adélie, Gentoo and Chinstrap) in the Antarcic peninsula sought to identify the prevalence of leucism in these different species. The paper cited in the margin provides the following information derived from detailed counts of penguin breeding colonies made during the years 1994-1997:

   Perform the following manipulations of the prevalence data for each species:

   (a) Express the prevalence as a fraction (a ratio of whole numbers).
   (b) Convert the prevalence to a decimal number.
   (c) Convert the decimal number to scientific notation.
   (d) Express the prevalence as a percentage of the population.
   (e) Express the prevalence in parts per million (ppm).
   (f) Determine the number of leucistic penguins in each count.

[14] Based on the article "Prevalence of leucism in Pygocelid penguins of the Antarctic Peninsula" by Forrest and Naveen, *Waterbirds* **23**(2): 283-285, 2000.

species	**prevalence**	**count**
Adélie	1:114,000	1,144,000
Chinstrap	1:146,000	293,800
Gentoo	1:20,000	41,550

3. From our discussion of standing timber values (Section 5.3.2), how would the result be different if we learned that the land parcel was 385 hectares instead of 1000?

6
Reasoning with Data

This chapter summarizes some of the key concepts and relationships of single-variable statistics that we might find useful for characterizing measurements, particularly when we have measured a quantity at multiple times, or we've measured many individual members of a population or collection. This is not intended to be an exhaustive introduction to statistics, and does not in any way substitute for a proper statistics course. It does, however, point to some connections that we can make between the measurement and characterization of data and the scientific description of nature that we sometimes seek.

6.1 Measurement and Sampling

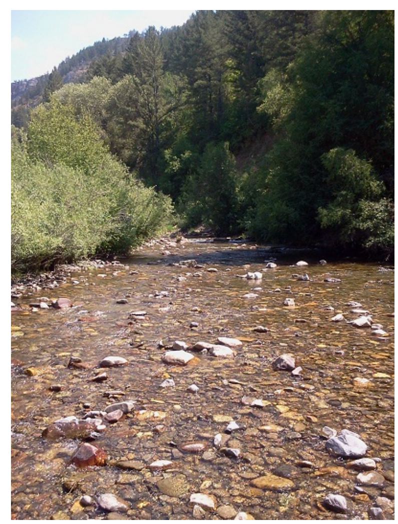


Figure 6.1: Cobbles on the bed of the Cub River, Idaho.

In the natural sciences we often need to estimate or measure a quantity or set of quantities that is too large, too numerous, or too complex to characterize completely in an efficient way. We can instead characterize it approximately with a *representative sample*. A representative sample is a small subset of the whole that is measured in order to characterize the whole.

Consider an example. In small headwater streams, many aspects of biotic health are linked with the size of the substrate – the sand, pebbles or boulders that compose the streambed. But it is impractical to measure all the gajillions of particles scattered over the entire bed. Instead, we attempt to get a smaller but representative sample of the bed material. This may be done in a number of different ways, but two common methods are: 1) to take one or more buckets full of sediment from the streambed and do a detailed particle-size analysis in a laboratory; and 2) measure the size of 100 randomly selected particles from the bed. Both methods obtain a sample, but each may represent the true streambed in a different way. The bucket method requires us to choose sample sites on the streambed. Our choices might be *biased* toward those places where sampling might be easier, the bed more visible, or the water shallower. In this case, our results might not be representative of the streambed as a whole.

The "pebble count" method, on the other hand, is intended to produce a more random sample of the streambed[1]. A person wading in the stream steps diagonally across the channel, and at each step places her index finger on the streambed immediately in front of the toe of her boot. The diameter of the particle that her finger touches first is measured, and then she repeats the process, zig-zagging across the channel until she has measured 100 (or some larger pre-determined number) particles. In principle, this *random sample* is more representative of the streambed, particularly as the number of particles in the sample is increased. Of course, increasing the number of particles in the sample increases the time and effort used, but with diminishing returns for improving the accuracy of the sample.

[1] This method is sometimes called the "Wolman pebble count" method for Reds Wolman, the scientist who first described and popularized it.

Hypothetically-speaking, an alternative pebble-count method could be to stretch a tape measure across the stream and measure the particle size at regular intervals, say every half meter. We can call this strategy the "point count" method. This alternative is appealing since it ensures that samples are distributed evenly across the channel and that samples are not clustered in space. However, it is conceivable that such *systematic sampling* could lead to a systematic bias[2]. If for example the streambed had clusters or patterns of particles in it that had a wavelength of 0.5 m, you could be inadvertently sampling only a certain part of the top of each dune, which might skew your results toward particle sizes that are concentrated on dune crests. Thus, a random sample is usually preferable as it is less susceptible to this kind of systematic bias.

[2] Systematic sampling is sometimes an easier, more straight-forward approach to sampling. However, if the setting within which sampling is taking place might have some systematic structure, systematic sampling could inadvertently bias the sample.

QUANTITIES DERIVED FROM A RANDOM SAMPLE are unrelated to one another in the same way that the size of one grain measured during a pebble count has no influence on the size of the next one. Part of our sequence of data might look like this:

> 12, 2, 5, 26, 4, 28, 19, 29, 3, 15, 31, 19, 24, 27, 7, 22, 28, 33, 21, 28, 13, 15, 25, 10, 14, 13, 16, 18, 33, 5

The random nature of this set of data allows us to use some of the familiar ways of describing our data, while boosting our confidence that we are also properly characterizing the larger system that we are sampling.

6.1.1 *Example: mark-recapture*

A frequent concern of the wildlife ecologist is the abundance and health of a particular species of interest. Ideally, we could count and assess the health of every individual in a population, but that is usually not practical - heck, we have a tough enough time counting and assessing the health of all the humans in a small town! Instead of trying to track down every individual though, we can do a decent job by simply taking a random sample from the population and performing the desired analysis on that random sample. As we have seen, if we are sufficiently careful about avoiding bias in our sampling, we can be reasonably confident that our sample will tell us something useful (and not misleading) about the larger population that the sample came from.

If our concern is mainly with the population of a target species in a certain area, we can use a method called *mark-recapture*, or *capture-recapture.* The basic premise is simple: we capture some number of individuals in a population at one time, band, tag or mark them in such a way that they can be recognized later as individuals that were previously captured, then release them. Some time later, after these individuals have dispersed into the population as a whole, we capture another set. The proportion of the individuals in the second capture who are marked should, in theory, be the same as the proportion of the whole population that we marked to begin with. If the number of individuals we marked the first time around is N_1, the number we captured the second time around is N_2, and the number in the second group that bore marks from the first capture is M, the population P may be estimated most simply as:

$$P = \frac{N_1 N_2}{M} \tag{6.1}$$

This comes from the assumption that our sample each time is random, and that the marked individuals have exactly the same likelihood of being in the second capture as they did in the first: $1/P$. Therefore, if we sampled and marked a fraction N_1/P the first time around and sample N_2 the second time around, then we should expect a fraction M/N_2 of them to be marked.

Of course this whole plan can be foiled if some key assumptions are not met. For example, we need the population to be "closed" – that is, individuals do not enter and leave the population such that our sample is not coming from the same set of individuals each time. Problems could also ensue if our "random" sample isn't random, if somehow the process of marking individuals either harmed them or made their likelihood of re-capture more or less likely, or if the time we allowed for them to re-mix with their population was not appropriate. On the last point, you can imagine that if we recapture tortoises 10 minutes after releasing them from their first capture, our second sample will not be very random. On the other hand, if we recapture marked fish 20 years after they were first marked, many of them may have died and been replaced by their offspring, and thus our assumption of a "closed" population is violated. So in planning a mark-recapture study, space and timescales need to be taken into account.

It is worth noting that the method described here is about the most stripped down version of mark-recapture. There are many modifications to the method and the equation used to compute population that either account for immigration/emigration, multiple recaptures, some possible re-recaptures, etc. There are also related methods using tagging and marking that can be used to explore the dispersal of individuals, migration routes and alot more!

6.2 *Describing measurements*

Measurements, or "data", can inform and influence much of a resource manager's work objectives, since they convey information about the systems of interest. Sometimes the data speak for themselves: raw numbers are sufficiently clear and compelling that nothing more needs to be done to let the data speak. More commonly, however, the data need to be summarized and characterized through one or more processes of **data processing** and **data reduction**. Processing might simply refer to a routine set of algorithms applied to raw data to make it satisfy the objectives of the project or problem. Data reduction usually summarizes a large set of data with a smaller set of descritptive statistics. For a set of measurements of a simple quantity, for example, we might wish to know:

Things we often want to know about our data

1. what is a typical observation?
2. how diverse are the data?
3. how should these properties of the data be characterized for different types of quantities?

The first point suggests the use of our measures of central tendency: mean, median and mode. The second goal relates to measures of spread or dispersion in the data. For example, how close are most values in the data set to the mean?

6.3 *Central tendency*

The central tendency of a data set is a characteristic central value that may be the **mean**, **median**, or **mode**. Which of these measures of central tendency best characterizes the data set depends on the nature of the data and what we wish to characterize about it.

Most of us are already familiar with the concept of a mean, or average value of a set of numbers. We normally just add together all of the observed values and divide by the number of values to get the mean. Actually, this is the *arithmetic mean*, and there are many alternative ways of computing different kinds of means that are useful in particular circumstances, but we won't worry about these now. For our purposes, the arithmetic mean is the mean we mean when we say mean or average. It would be mean to say otherwise.

Before continuing, lets briefly discuss the different kinds of notation what we might use when talking about data. To define something like the mean with an equation, we'd like to make the definition as general as possible, i.e., applicable to all cases rather than just one. So we need notation that, for example, does not specify the number of data points in the data set but allows that to vary. If we want to find the mean (call it $\bar{x}$) of a set of 6 data points (x_1, x_2, and so on), one correct formula might look like this:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6}{6} \tag{6.2}$$

and of course this is correct. But we can't use the same formula for a dataset that has 7 or 8 values, or anything other than 6 values. Furthermore, it is not very convenient to have to write out each term in the numerator if the data set is really large. So we need a shorthand that is both brief and not specific to a certain number of data points. One approach is to write:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_n}{n} \tag{6.3}$$

where we understand that n is the number of observations in the data set. The ellipsis in the numerator denotes all the missing values between x_2 and x_n, the last value to be included in the average. Using this type of equation to define the mean is much more general than the first example, and is more compact as long as there are 4 or more values to be averaged.

One additional way you might see the mean defined is using so-called "sigma notation"[3], where it looks like this:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{6.4}$$

where the big Σ is the summation symbol. If you've never encountered this before, here's how to interpret it: the "summand", the stuff after the Σ, is to be interpreted as a list of values (in this case x_i) that need to be added together, and i starts at 1 and increases until you get to n. You can see the rules for what i means by looking at the text below and above the Σ. Below where it says $i = 1$ that means

[3] This symbol is a handy shorthand for the process of adding a bunch of quantities together, but also serves the purpose of scaring many poor students away. Once you realize that it's just an abbreviation for listing all the the terms to be added ($x_1 + x_2 + ...$) and some of the rules for doing so, it becomes a tad less fearsome.

that i begins with a value of 1 and increases with each added term until $i = n$, which is the last term. So in the end, you can interpret this to have a meaning identical to the equivalent expressions above, but in some cases this notation can be more compact and explicit. It also looks fancier and more intimidating, so people will sometimes use this notation to scare you off, even though it gives you the same result as the second equation above.

6.3.1 *Mean versus Median*

For some data sets, the mean can be a misleading way to describe the central tendency. If your creel after a day of fishing includes 5 half-pound crappies, a 3/4-pound walleye, 4 one-pound smallmouths and one 16-pound muskie, it would be correct but misleading to say that the average size of the fish you caught was 2.1 pounds. The distribution of weights includes one distant outlier, the muskie, that greatly distorts the mean, but all of the other fish you caught weighed one pound or less. We might say in this case that the mean is sensitive to outliers.

species	weight (lbs.)
crappie	0.5
crappie	0.5
crappie	0.5
crappie	0.5
crappie	0.5
walleye	0.75
smallmouth	1.0
smallmouth	1.0
smallmouth	1.0
smallmouth	1.0
muskie	16
mean	2.1
median	0.75

Table 6.1: A decent day's catch on the lake.

The median is an alternative measure of central tendency that is not sensitive to outliers. It is simply the value for which half the observations are greater and half are smaller. From your fishing catch, the 0.75 pound walleye represents the median value, since 5 fish (the crappies) were smaller and 5 fish (the smallies and the muskie) were larger. The median may also be thought of as the middle value in a sorted list of values, although there is really only a distinct middle value when you have an odd number of observations. In the event that you've got an even number of observations, the median is halfway between the two middle observations.

6.3.2 *Mode*

The mode is the value or range of values that occurs most frequently in a data set. Since you caught 5 half-pound fish and fewer of every other weight value in the dataset, the mode of this distribution is 0.5 pounds. Now if the weights we've reported above are actually rounded from true measured weights that differ slightly, this definition becomes less satisfactory. For example, suppose the half-pound crappies actually weighed 0.46, 0.49, 0.5, 0.55 and 0.61 pounds. None of these are actually the same value, so can we say that this is still a mode? Indeed we can if we choose to discretize or *bin* these data. We might say that our fish weights fall into bins that range from 0.375 to 0.625, 0.625 to 0.875, 0.875 to 1.125, and so on. In this case, since all of our crappies fall in the range 0.375 to 0.625 (which is $5 \pm 1/8$ lbs), this size range remains the mode of the data set. We can see this visually in a histogram, which is just a bar-chart showing how often measurements fall within each bin in a range (Figure 6.2).

It is permissible to identify multiple modes in a data set if it improves the description. The first mode is the data bin that appears most frequently, but second and third and additional modes can be used as well. A second mode in our fish sample is in the 1-pound bin, which included 4 smallmouth bass. It is particularly useful in the case of *multimodal* data sets to report the modes because the multimodal nature of the data set cannot be represented by the mean or the median. In fact, if you were only presented with the list of weights, you might still have a hunch that there were multiple species or multiple age-classes present in the creel due to the multimodal weights.

In practice, reporting all of these measures of central tendency may deliver the most complete picture of data, but as we've seen each is particularly useful in some cases and can be misleading in others. That said, we can actually infer additional properties of the dataset by noting, for example, the difference between the mean and median.

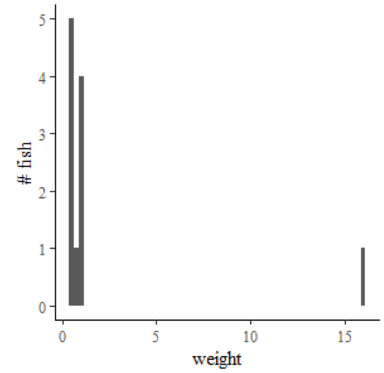


Figure 6.2: A histogram showing the frequency of observations of fish weights. The height of each bar corresponds to the number of fish in each of the weight bins along the horizontal axis. Values are scrunched to the edges because the large dispersion of data.

In some cases, multi-modal data can be suggestive of a mixed sample; that is, there is more than just one type of thing or from more than just one source in the sample.

6.4 *Spread*

As mentioned previously, one way to quantify dispersion of a data set is to find the difference between any given observation and the expected value or sample mean. If we write this:

$$x_i - \bar{x}, \tag{6.5}$$

we can call each such difference a **residual**. A could be used to describe the relationship between individual data points and the sample mean, but doesn't by itself characterize the spread of the entire data set. But what if we add together all of these residuals and divide by the number of data points? Well, this should just give us zero, according to the definition of the mean! But suppose instead that we *squared* the residuals before adding them together. The formula would look like:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{6.6}$$

This expression is defined as the **variance** and is strangely denoted by σ^2, but you'll see why in a minute. Squaring the residuals made most of them larger and made negative residuals positive. It also accentuated those outlier data points that were farther from the mean. Now if we take the square root of the variance, we're left with a finite positive value that very well represents how far data typically are from the mean: the **standard deviation** of the sample, or σ^4. The

[4] If you keep track of the units of these different measures of spread, you'll notice that the standard deviation should have the same units that the original data, x_i does.

formal definition of standard deviation looks like this:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{6.7}$$

The gives us a good sense for how far from the mean a typical measurement lies. We can now characterize a sample as having a mean value of $\bar{x}$ and standard deviation of σ, or saying that typical values are $\bar{x} \pm \sigma$. But in reality, if we computed $\bar{x}$ and σ, the bounds set by $\bar{x} - \sigma$ and $\bar{x} + \sigma$ only contain about 68% of the data points. If we want to include more of the data, we could use two standard deviations above and below the mean, in which case we've bounded more than 95% of the data.

6.5 *Error & Uncertainty*

One piece of information we have thus far omitted from our list of properties that fully define a quantity's value is uncertainty. This is particularly important when we are quantifying something that has been measured directly or derived from measurements. Thus, to even more completely define the value of a *measured quantity*, we should include some estimate of the uncertainty associated with the number assigned to it. This will often look like:

$$x = x_{best} \pm \delta x, \tag{6.8}$$

where x is the thing we are trying to quantify, x_{best} is our best guess of its value, and δx is our estimate of the uncertainty. Though it will depend on the quantity in question, our best estimate will often be the result of a single measurement or – better yet – the mean of a number of repeated measurements.

The preferred value x_{best} for a quantity of interest will often be the **mean** of repeated measurements of that quantity.

6.5.1 *Uncertainty in measured quantities*

All measurements are subject to some degree of uncertainty, arising from the limited resolution of the instrument or scale used to make the measurement, or from random or systematic errors resulting from the method or circumstances of measurement. Let's consider an example:

Suppose two fisheries biologists each measured the lengths of ten of the brook trout captured during the electrofishing traverse from Problem 3.7. Both used boards with identical scales printed on them, graduated to half of a centimeter. They then plan to put their measurements together to get a data set of 20 fish. One of them was trained to pinch together the tail fins to make this measurement, while the other was not. In addition, because they wished not to harm the fish, they made their measurements quickly, even if the fish flopped and wiggled during the measurement. What are the potential sources of error and how big are they relative to one another?

For starters, implicit in the graduations on this board is that the user cannot confidently read any better than half-centimeters off the scale. He or she can, however, visually *interpolate* between two adjacent graduations to improve precision (see below). However, this step is inherently subjective and limits the certainty of the measurement. We might call this **instrumental error** because its magnitude is set by the instrument or device use to make the measurement. One way to reduce this source of error is to use a more finely-graduated scale.

Instrumental error is fixed by the resolution of the device used to make a measurement, and can usually only be reduced by using a more precise instrument.

A second source of error arises from the hasty measurements and the fact that the fish were not necessarily cooperative. Perhaps the mouth was sometimes not pressed up all the way against the stop, or the fish wasn't well aligned with the scale. Some lengths may have been too large or small as a result, yielding a source of error that was essentially random. Indeed, we can call this **random error** since its sign and magnitude are largely unrelated from one measurement to the next. Reducing this source of error in this case would require either more careful and deliberate effort at aligning and immobilizing the fish, or making multiple measurements of the same fish. Both of these solutions could endanger the fish and may therefore not be desirable.

Random measurement errors may be mitigated by repeating measurements.

A third source of error is associated with the difference in the way the two scientists dealt with the tail fin. Length measurements made with the fins pinched together will usually be longer than those without. Had they measured the same group of ten fish, one set of measurements would have yielded lengths consistently smaller than the other. This is a **systematic error**, and can often be troublesome and difficult to detect. This highlights the need for a procedural statement that establishes clear guidelines for measurements wherever such sources of systematic error can arise.

Systematic errors result in data that deviate systematically from the true values. These errors may often be more difficult to detect and correct, and data collection efforts should make great pains to eliminate any sources of systematic error.

Each of these types of error can affect the results of the measurements, and should be quantified and included in the description of the best estimate of fish length. But errors can affect the best estimate in different ways. Instrumental error, as described above, can itself either be random or systematic. The printed scale on one of the fish measurement boards could be stretched by a factor of 3% compared to the other, resulting in a systematic error. Likewise one board might be made from plastic that is more slippery than the other and thus more difficult to align the fish on. This could result in additional random error associated with that device. But what are the relationships between these types of errors and the best estimate that we are seeking?

Error or variation? Questions to ask yourself

1. What were possible sources of error in your measurements? Are they random or systematic?

2. How can you tell the difference between error in measurement and natural variability?

6.5.2 *Real variability*

Not all deviations from the mean are errors. For real quantities in nature, there is no good reason to assume that, for example, all age-0 brook trout will be the same length. Indeed we expect that there are real variations among fish of a single age cohort due to differences in genetics, feeding patterns, and other real factors. If we're measuring a group of age-0 fish to get a handle on how those fish vary in size, then at least some of the variation in our data reflects real variation in the length of those fish. How do we tease out the variation that is due to errors from the variation that is due to real variability?

Often a good approach is to try to independently estimate the magnitude of the measurement errors. If those measurement errors are about the same magnitude as the variations (residuals) within the data, then it may not be possible to identify real variability. However, in the more likely event that our measurements are reasonably accurate and have small measurement errors compared to their spread about the mean, then the indicated variations probably reflect true variability.

This observation returns us to our earlier question: when we seek to characterize some quantity how should we identify our best estimate and our degree of uncertainty in that estimate. If we wish to characterize a single quantity and our certainty that our best estimate is close to or equal to the true value, we should use the mean of repeated measurements of this value and the standard error of those measurements. The standard error can be readily estimated by dividing the standard deviation of the repeated measures by the number of measurements n:

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}}. \tag{6.9}$$

This should be equivalent to the standard deviation of a number of estimates of the mean $\bar{x}$, if several samples were taken from the full population of measurements. Like the standard deviation, we can be about about 68% confident that the range $x_{best} + \mathrm{SE}$ to $x_{best} - \mathrm{SE}$ includes the true value we wish to characterize, but if we use 1.96 SE instead, we can have 95% confidence[5]. A complete statement, then, of

[5] Note that we are currently assuming that our measurements are normally distributed.

our best estimate with 95% certainty in this context is to say:

$$x = x_{best} \pm 1.96 \text{ SE}, \tag{6.10}$$

If instead we desire a characterization of a typical value and range for something that has real variability among individuals in a population, we will usually describe it with the mean and standard deviation.

$$x = x_{best} \pm 1.96\ \sigma, \tag{6.11}$$

6.6 *Distributions*

The kind of data we've been talking about thus far is univariate: a single quantity with variable values like the diameter of a streambed particle, or the length of a fish. As we know, not all age-0 brook trout are the same size. In a first-pass capture of 50 fish, for example, we should expect some variability in length that might reflect age, genetics, social structure, or any other factor that might influence development. The variation may be visualized graphically in a number of ways. We'll start with a histogram.

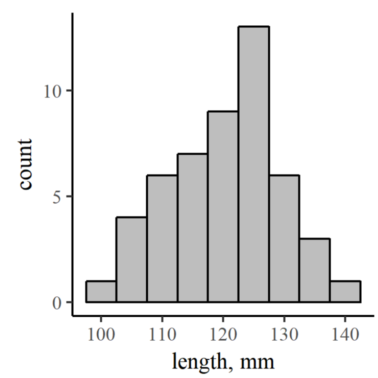


Figure 6.3: A histogram showing the distribution of simulated (random) measurements of the length of 100 snakes.

A histogram shows the distribution of a set of *discrete* measurements – that is the range of values and the number of data points falling into each of a number of bins, which are just ranges of values (112.5 to 117.5 is one bin, 117.5 to 122.5 another...). This can be called a frequency distribution, and a histogram is one of the best ways to visualize a frequency distribution (Figure 6.3).

But what if we had uniformly distributed data? A uniform distribution means that it is equally likely that we'll find an individual with a length on the low end (97.5-102.5 mm) of the range as any other. That would look quite different – there would be no hump in the middle of the histogram, but rather a similar number of measurements of each possible length. The uniform distribution is great: in fact, we count on uniformity sometimes. If you are at the casino and rolling the dice, you probably assume (unless you're dishonest) that there is an equal probability that you'll roll a 6 as there is that you'll roll a 1 on any given die. We can call that a uniform probability distribution for a single roll of a die. But What if the game you are playing counts the sum of the numbers on 5 dice? Is there still a uniform probability of getting any total value from 5 to 30?

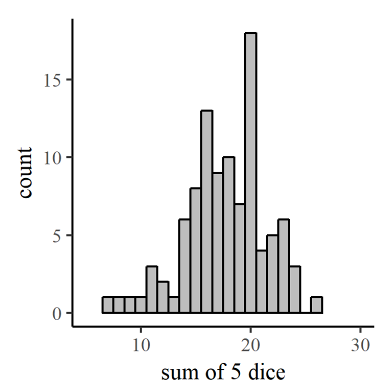


Figure 6.4: Sum of the values of five dice, rolled 100 times each.

We could actually simulate that pretty easily by randomly choosing (with a computer program like R[6] or Excel) five integers between 1 and 6 and adding them together. Figure 6.4 shows the plot that comes out. Looks sorta like a bell curve, right? Well, how likely is it that you'll get five 1's or five 6's? Not very, right? You're no more likely to get one each of 1,2,3,4 and 5 either, right? However, there

[6] R is a top choice software for general-purpose data analysis and modeling. It is free software, works on most computer platforms, and has nearly infinite capabilities due to the user-contributed package repository. Learn more about R at https://cran.r-project.org/

are multiple ways to get a 1,2,3,4 and 5 with different dice showing each of the possible numbers, whereas there is only one way to get all sixes and one way to get all ones. So there are better chances that you'll get a *random* assortment of numbers, some higher and some lower, and their sum will tend toward a central value, the mean of the possible values. So, since your collection of rolls of the dice represent a random sample from a uniform distribution, the sum of several rolls will be normally distributed.

What's it got to do with fish? If we sample brook trout randomly from one stream reach and measure their lengths, we might expect them to be normally distributed. Describing such a normal distribution with quantities like the mean and standard deviation gives us the power to compare different populations, or to decide whether some individuals are outliers. The nuts and bolts of those comparisons depend on how the type of distribution represented by the population. An ideal normal distribution is defined by this equation:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{6.12}$$

and it's graph, in the context of our original hypothetical distribution of fish lengths, looks like the red line in Figure 6.5. In order to compare the continuous and discrete distributions, we've divided the counts in each bin by the total number in the sample (50), to yield a *density* distribution. The blue line is just a smoothed interpolation of the top centers of each bar in the discrete distribution, so it generally reflects the density of data within each bin. As you can see, the discrete distribution density and the continuous normal distribution functions are similar, but there are some bumps in the discrete distribution that don't quite match the continuous curve. As you can imagine though, that difference would become less pronounced as your dataset grows larger. Related to this, then, is the idea that your *confidence* in the central tendency and spread derived from your dataset should get better with more data.

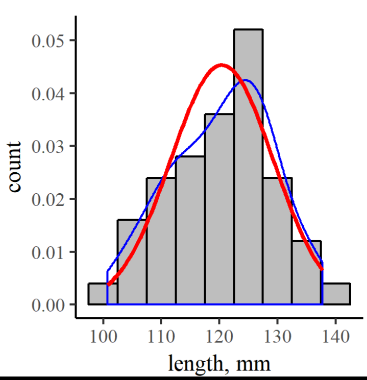


Figure 6.5: Superimposed discrete distribution density (bars), interpolated continuous density from the discrete distribution (blue line), and an ideal continuous distribution function with the same mean and standard deviation.

Exercises

1. Download the data from Derek Ogle's InchLake2 dataset from the fishR data website. Using either a spreadsheet or data analysis package, isolate the bluegill from the dataset and identify the following:

   (a) Mean bluegill length.

(b) Standard deviation of bluegill length.

(c) Mean bluegill weight.

(d) Standard deviation of bluegill weight.

2. The graph and data table below and right show measurements of brook trout lengths from pass #1 of the electrofishing campaign described in Problem 3.7. Use these resources to answer the following questions:

   (a) Judging from the histogram in Figure 2, does the dataset contain just one mode or more than one? What might be the reason for this?

   (b) What is the mean and standard-deviation for the (presumed) age-0 portion of this sample?

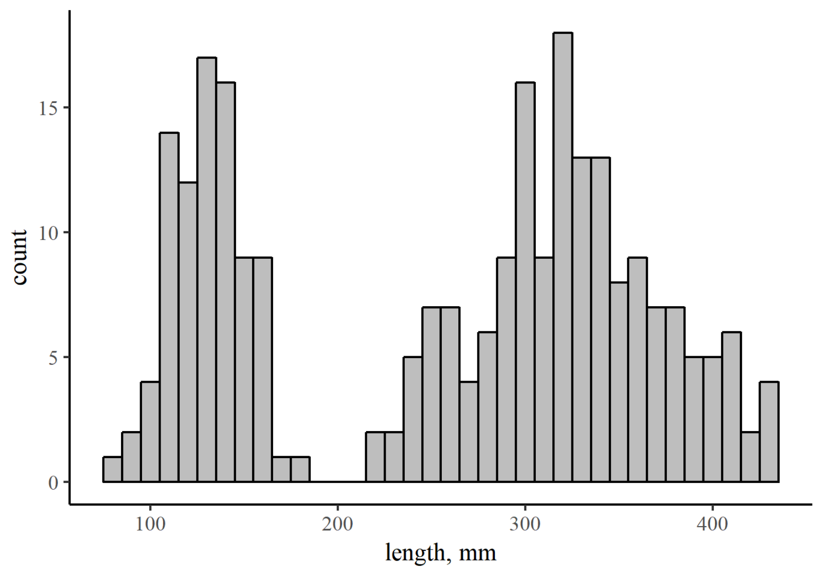


Figure 6.6: A complete frequency distribution of brook trout lengths from electrofishing pass #1 from Problem 3.7.

index	1	2	3	4	5	6	7	8	9	10
1	313	135	342	297	137	112	379	116	142	154
2	288	322	241	364	360	348	265	127	297	143
3	355	110	152	107	157	338	135	345	251	110
4	127	372	164	417	364	358	113	329	83	366
5	305	343	129	378	298	245	392	121	371	394
6	256	397	114	292	146	147	243	320	294	154
7	406	301	156	294	396	132	296	349	247	313
8	261	406	332	381	329	250	233	316	130	104
9	248	294	427	295	316	339	328	255	344	121
10	312	339	271	323	272	259	120	123	316	301
11	401	114	279	160	293	321	217	301	240	133
12	135	370	275	137	139	130	276	299	296	111
13	323	250	414	308	317	362	336	332	429	114
14	141	163	264	325	151	167	380	100	138	120
15	160	321	246	351	369	146	284	108	131	136
16	263	131	376	374	419	310	431	121	321	326
17	125	410	312	347	113	297	89	96	294	134
18	342	356	110	131	139	296	285	99	313	372
19	361	428	344	301	365	347	283	158	331	397
20	149	155	307	165	321	224	137	333	132	231
21	329	133	305	388	319	120	389	330	411	143
22	306	261	359	126	143	386	338	179	319	140
23	273	320	122	144	384	112	408	316	344	303
24	122	324	137	331	92	113	341	399	353	305
25	287	117	354	332	376	282	244	335	157	144

Table 6.2: Brook trout length data from electrofishing pass #1. All lengths in mm.

7 Interlude: Collecting and managing data

Data is information. Data is the result of somebody's efforts to record and store information, often to provide an opportunity for insight. It can be used to discover patterns, test hypotheses, and support arguments, among other things. But the numbers themselves often cannot convey much meaning – it is through manipulation and interpretation of the data that those uses can be realized. We therefore need to be conscious about how data are structured and managed, so that they can be manipulated and interpreted to reveal insights. If data are badly structured and managed, the risks can be great. Optimistically, we risk wasting lots of time trying to restructure data to allow the kind of analysis we wish to perform; worse yet, we risk compromising the integrity of or, heaven-forbid, the complete loss of data through poor structuring and mismanagement.

There are some great resources for data management out there in various forms, including some geared toward biologists. A few great ones are:
Data Carpentry
www.datacarpentry.org.
Wickham, H., 2014, Tidy Data. *Journal of Statistical Software* 59(10).
Saltz, J.H. and J.M. Stanton, 2017, *Introduction to Data Science*, Sage publ.

First, lets be clear about what is meant by data structure and data management:

- **Data structure** refers to the organization and layout of data as it is stored. Whether it is handwritten in notebooks or stored in spreadsheets or text files, data usually has an architecture that reflects the intentions (or ignorance) of the data manager or sometimes the protocols of his/her organization or institution.
- **Data management** is the set of practices aimed at preserving the quality, integrity, and accessibility of the data. This can include all phases of data usage from collection and manipulation to storage, sharing, and archiving.

7.1 *Who is data for?*

Unless you work with highly classified or proprietary information and are required to protect and encode your data, you likely need data to be readily understood and usable, not just to you but to others you work with or the broader public[1]. But we also need to realize

[1] Many government funding agencies such as the National Science Foundation, US Department of Agriculture, and National Institutes of Health now require that researchers develop a data management plan that includes strategies for structuring, archiving and indexing data in publicly-accessible repositories.

that the humans who need to make sense of the data will be using tools like computers to facilitate this approach. Therefore the data structure needs to also accommodate the demands of the computer hardware and software that it is used on as well as the humans. Thus, data should be organized in a logical, self-consistent way and it should be accompanied by documentation that helps to explain the content and and context of the data. Similarly, accessible data archiving, in principle, allows colleagues and competitors to test, verify, reproduce, and/or compare results with their own, ensuring that scientific advances that you make with the help of your data can also lead to advancement of science and management more broadly.

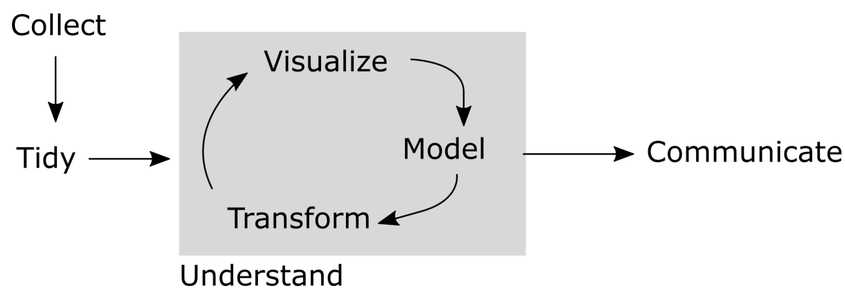


Figure 7.1: A typical workflow for data. After Grolemund and Wickham, 2017, *R for Data Science*, O'Reilly.

Consider the schematic workflow illustrated in Figure 7.1. Once collected, data must be organized and formatted in a way that facilitates their analysis on a computer. A popular term for data that are formatted to simplify computer manipulation is **tidy data** (more on this below). When this process is complete, the data may be analyzed as needed to address the problem or hypothesis at hand. This process of making sense of the data may then produce a result that needs to be communicated back to humans. When data is presented for consumption by the human eye and brain, the organization and structure should reflect the expectations and attention span of the humans. Unless the data set is small, the raw or transformed data may not be appropriate to display. Instead, summary data are more appropriate, either in narrative, table, or graphical format.

7.2 Tidy data

To understand the significance of tidyness, it is perhaps helpful to consider untidy or kludgy data. Below is a portion of a data table containing the weights and lengths of small fish captured during a population survey of Inch Lake, Wisconsin[2]. Let's unpack this data set. There are two different fish species listed, one observed both in 2007 and 2008, and one only in 2007. Lengths and widths are provided for all fish measured, but there are different numbers of

[2] This is a deliberately disorganized snippet based on a dataset from Derek Ogle's neat website http://www.derekogle.com/fishR/data

Table 7.1: A portion of an untidy dataset.

number	bluegill (2008)	bluegill (2007)	Iowa Darter (2007)
1	L:1.5 W:0.7	L:1.9 W:1.3	L:2.1 W:0.9
2	L:1.0 W:0.7	L:1.6 W:1.3	L:2.0 W:1.3
3	L:2.6 W:1.5	L:2.4 W:1.7	L:1.7 W:0.7
4		L:1.1 W:0.6	

fish in each column. There is also an index value in the first column that facilitates counting the number of fish of each species captured in each year. This is reasonably straight-forward for a human to interpret, particularly if we are told that L corresponds to a length in inches and W corresponds to weight in grams. However a larger dataset organized like this table would be miserable to analyze for a variety of reasons, including:

- columns don't have the same number of values
- the same species has data across multiple columns
- two variables (length and width) are listed together within each column, with numbers and letters mixed

One instructive question to ask is how many variables there are here. We notice that data span multiple years, so year could be a variable. There are also multiple species here, so species could be viewed as a variable. Then length and width should each be variables. Finally, if we wish to have an index or ID number for each fish, that might be a fifth variable[3]. In general, tidy data is organized in a rectangular array in which each column represents a variable and each row an observation. In most cases, the first row contains descriptive but simple column headers. This simple prescription seems unthreatening, but it is often surprising how pervasive untidy data is.

[3] If there are multiple datasets derived from the same group of fish, assigning a fish ID number would be a simple way to connect these datasets using database methods.

Tidy data has:

- one column for each variable
- one row for each observation
- a header row

So with five variables, how many observations do we have? Each fish represents an observation, with a unique ID number, year of capture, species, weight and length. From Table 7.2, there appear to be ten fish listed among the three columns. so since each fish is an observation. According to the principles of tidy data, then, there should be ten rows of data with values in each of the five columns. Below is a tidy representation of this dataset. This table is now organized in a way that can easily be sorted, filtered, and summarized in common statistics and computational software packages.

7.3 *Data management*

Because data is often the hard-won result of costly and time-consuming observations and measurements, its management should be deliber-

fishID	year	species	weight (g)	length (in)
1	2008	bluegill	1.5	0.7
2	2007	bluegill	1.9	1.3
3	2007	Iowa darter	2.1	0.9
4	2008	bluegill	1.0	0.7
5	2007	bluegill	1.6	1.3
6	2007	Iowa darter	2.0	1.3
7	2008	bluegill	2.6	1.5
8	2007	bluegill	2.4	1.7
9	2007	Iowa darter	1.7	0.7
10	2007	bluegill	1.1	0.6

Table 7.2: The data from Table 7.2 transformed to a tidy dataset.

ate and careful. Well-managed data can be stored, retrieved, analyzed, and used to develop insights or aid in management decisions without compromising the data itself, and without spending unnecessary time and energy in decoding and interpreting the raw data. Thus, proper management entails not only careful structure as described above, but also well-organized storage and full documentation.

First, important raw data should be stored redundantly. If it's only in hard-copy (e.g., in field notebooks), consider scanning or transcribing the hard-copy data to preserve a digital version that can be backed-up regularly. When the data results from original research that can be shared publically, it can be uploaded to data repositories[4] A complete set of data should be archived and never modified, while data reduction and analysis are done on copies of the formatted raw data.

[4] Data repositories like the LTER Data Portal require formatted data and metadata to ensure long-term accessibility and adequate documentation.

When analysis or reduction occurs much later than the time of collection and storage, or is done by a different person or group than the researcher who collected the data, adequate documentation or metadata is essential. Metadata can include narrative descriptions of where the data was collected, when and how it was collected, and should include references or links to any published or publically-available research or information stemming from the data. The metadata should always include a data dictionary, fully describing the quantities represented by each of the variables collected (i.e., variable name, symbol, units, and procedural statement). These guidelines ensure that data remain safe, useful, and accessible.

Part III

SPATIAL REASONING

8
Geometry and Geography

One of the fundamental types of quantities that we use as natural resource scientists and professionals is a measure of distance or size. Whether we're describing the board-feet of merchantable lumber in a ponderosa pine, the fork length of a trout, the size of a white-tail deer's home range, or the storage capacity of a flood-control reservoir, we are concerned with spatial quantities that ultimately manifest from linear measurements in space. Much of this may feel familiar to you, but there are important messages to take home from working with both simple and compound spatial quantities that will serve you well in working with maps, photos, design plans, and other tools that professionals use.

8.1 Length, Area, and Volume

Consider the wetland shown in the map below (Figure 8.1). How might we characterize its size? Perhaps the answer depends on the context of the question. Are we interested in how far it is to cross it in a canoe? How long the shoreline habitat is for waterfowl? The number of acres it occupies? What about the amount of water it holds? In turn these questions point to linear distance, curvilinear distance, area and volume, respectively. Each of these types of quantities can be expressed in a variety of ways[1] according to the setting, the application, or the purpose of communication.

[1] For example, we can express all the quantities in terms of SI units of m, m^2, and m^3, or we can use more traditional U.S. agricultural units like feet, acres, and acre-feet.

The distinctions between length, area, and volume are more than just trivia. They clearly reflect not only different ways of estimating size, but different numbers of spatial dimensions. In common parlance, length is one-dimensional or 1D, area is two-dimensional or 2D, and volume is three-dimensional or 3D. The units can be a clue to how many dimensions are indicated in a spatial description. From Chapter 4, recall that we can generalize units in terms of the fundamental dimensions they entail. From this perspective then, a 1D length has dimensions of $[L]$, a 2D area $[L^2]$, and a 3D volume

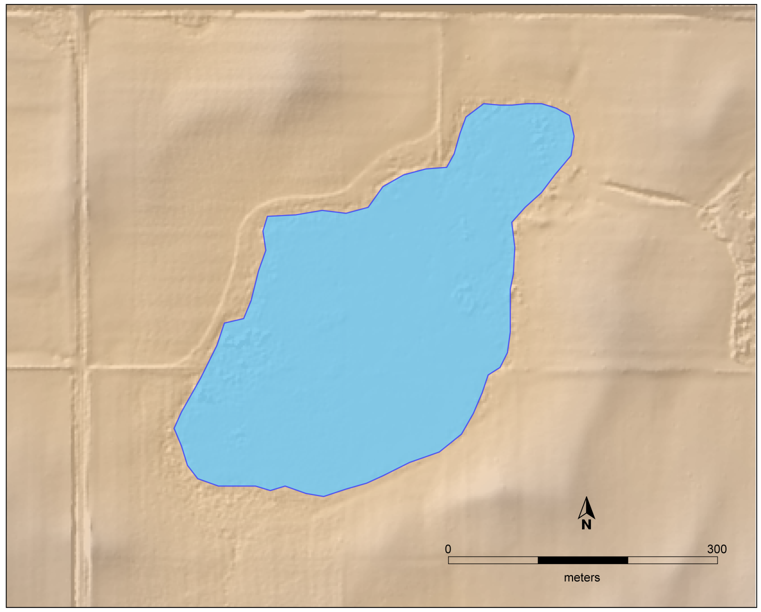


Figure 8.1: Map of a wetland in Iowa. The blue polygon shows the extent of seasonal open water overlain on shaded relief.

$[L^3]$. This is true regardless of the specific units used to describe the quantity of interest, though sometimes the dimensionality can be obscured by the use of compound units. For example, if we're told that a woodlot is 820,000 ft^2, that is a straight-forward 2D measure of area. If that same woodlot is described as 126,000 board feet[2], now we're talking about a *volume* of wood expressed in a unit that is not particularly transparent to outsiders, though it is customary among foresters. If we need to do computations involving quantities of this sort we need to be certain that we understand what makes sense and what doesn't make sense to do; what's permissible and what isn't.

[2] One board-foot is equivalent to 144 cubic inches of merchantable timber; can be visualized as a 12 inch long and 12 inch wide board that is 1 inch thick.

Note that some of our commonly-used spatial units are compound by definition. An acre, for example, is a unit of area even though it is not expressed in a squared-length form. Originally defined as the area of land that could be plowed with oxen in a day, an acre is 43,560 ft^2 or roughly 4,047 m^2. If you measured out a square 208.7 ft on a side, that would be approximately an acre. The acre's cousin in SI units, the hectare, is also a unit of area with a simpler definition: an area of land 100 m wide and 100 m long, or 10,000 m^2.[3]

[3] Neither the acre nor the hectare need be any particular shape, nor do they necessarily need to be contiguous, though they usually are.

We're also familiar with several alternative ways of expressing volume with derived units, particularly when talking about liquids. It's not unheard-of to talk about fluid volumes in cubic meters or cubic feet (particularly if we're referring to volume per unit time as we do when describing river flows in cubic feet per secod or cfs), but it is more common to hear fluid volumes expressed in gallons, millileters or liters. These are all legitimate expressions of fluid (gas or liquid) volumes and some have relatively straight-forward relationships to length-cubed volumes: for example, 1 ml is the same as 1 cm^3 or cc. However, if we wish to perform computations more complex than addition or subtraction on such quantities, it can be advantageous to convert them into more fundamental units like cubic meters. One interesting unit of volume mentioned above is the acre-foot, which (as you might guess) is the volume corresponding to a one acre area of something that is one foot deep. This means its dimensions are an area $[L^2]$ times a depth $[L]$, therefore its a volume $[L^3]$. We encounter this unit of volume sometimes in descriptions of ponds or stormwater-basins because it may be easier to visualize, but this can also make it more difficult to perform computations.

8.1.1 Unit conversions in space

Here's an common exercise that American students often need to perform in earth science, geography or natural resource courses[4]: measure a rectangular land area on a USGS map with a scale in feet and miles, and convert it to square meters or square kilometers.

[4] perhaps this exercise is going the way of the paper map itself as people increasingly interact with only digital maps these days!

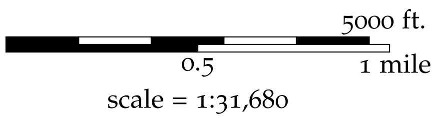


Figure 8.2: A generic map scalebar showing map distances in feet and miles.

This exercise will typically begin with each student begrudgingly making tickmarks on the edge of a sheet of paper that is lined up with the map scale. For purposes of illustration, we'll follow the hypothetical (but not uncommon) path of a student who is prone to making some common mistakes. Our student uses the marked paper to estimate the length of each side of the rectangular land area using the scaled map units[5]. Perhaps the values are 6.2 miles and 2.1 miles. He then proceeds to multiply them together, because he's aware that the area of a rectangle is the product of its sides. So he punches 6.2×2.1 into his calculator and gets 13.02. When asked to supply units for his answer, he reasons that since the units he was measuring distances in were miles, the answer is also in miles. His instructor points out that miles are a unit of length not area, and that he should write the equation out complete with units to ensure that his result comes out in units of area. So he writes:

[5] see the next section for more on scaling

COMMON MISTAKE #1: leaving the units out of computations can lead to errors in unit assignment for solutions.

$$6.2 \text{ mi} \times 2.1 \text{ mi} = 13.02 \text{ mi}^2$$

and the instructor nods approval but says, "and now we need the area in square kilometers". Our downtrodden student proceeds to look up the conversion factor between miles and kilometers: $1\text{mi} \simeq 1.609$ km. Great. The calculator buttons click away until the student, exasperated, inquires "so the area in square km is 13.02×1.609, which equals about 20.95 right?" The ever-patient instructor shakes her head: "there are 1.609 km in a mile, but how many square km in a square mile?". Our student stares at the map and feigns interest in the question. On a whim, he asks "do I need to square the 1.609 too?" The instructor pats him on the shoulder and remarks "yep, write it all out, and don't forget the units" as she walks away. Our student, relieved at having guessed correctly, writes:

COMMON MISTAKE #2: using length conversion factor for area conversion.

$$6.2 \text{ mi} \times 2.1 \text{ mi} = 13.02 \text{ mi}^2$$

$$13.02 \text{ mi}^2 \times \left(1.609 \frac{\text{km}}{\text{mi}}\right)^2 = 33.71 \text{ km}^2$$

To see why we need to square the conversion factor like our student ultimately did, let's write out the conversion equation the way he did it at first, but using only the units (this is a variation of the dimensional homogeneity heuristic from Chapter 4):

$$\text{mi}^2 \times \frac{\text{km}}{\text{mi}} = \text{km}^2$$

If we cancel common units, we *should* be able to show that the units of the left-hand side are equal to the units on the right-hand side, but here we can only cancel miles in the numerator from miles in the denominator on the left-hand side, leaving a meaningless unit equation: mi km = km^2. That can't be true.

If instead we reason that our conversion factor needs to allow us to cancel through to make the units equivalent on both sides, we square the conversion factor and its units to yield:

HEURISTIC: Unit conversions can be written as equations with the current quantity and units on the left-hand side and the quantity in desired units on the right-hand side. All conversion factors should include units and be subject to operations such that the expression satisfies unit and dimensional homogeneity.

$$\text{mi}^2 \times \left(\frac{\text{km}}{\text{mi}}\right)^2 = \text{km}^2$$

This approach can be generalized for other types of spatial unit conversions as well, provided that our original and desired units are not compound. Suppose we are measuring the size of something in units based on the length unit U_1 and we need to convert it into units based on the length unit U_2. If the conversion factor between U_1 and U_2 is $C_{1\to2}$, the conversion equation can be written:

$$U_1^d \times C_{1\to2}^d = U_2^d \tag{8.1}$$

In this equation, d is the spatial dimensionality of the quantity, so its 1 for lengths, 2 for areas, and 3 for volumes. It's important to note that the conversion factor $C_{1\to2}$ should correspond to the number of unit 2's per unit 1, as we did for the map area conversion above.

Note that unit conversion factors between compound units like acres and hectares are *not* subject to these concerns. There is no such thing as a square acre because an acre is already a unit of area, so nothing needs to be done to the conversion factor if you are converting from, for example, acres to hectares: there are 0.4047 hectares in an acre, period. In this way these compound units can make life easier, but if you are simultaneously working with other quantities in meters, this convenience comes at a price.

IT MIGHT HAVE OCCURRED TO YOU to approach the map problem in a slightly different way. Suppose that instead of computing the area in square miles immediately after measuring the sides of the rectangle, our student had converted the sides from miles to kilometers first. Does this make any difference?

In this case, the map distances are $6.2 \times 1.609 = 9.976$km and $2.1 \times 1.609 = 3.379$km. Their product is 33.71km^2, which is the same result as before. That should come as no surprise, since the only difference is that the unit conversions from miles to km occurred before finding the area rather than after. Indeed this is one way to make the problem a bit simpler to think about, but thanks to the commutative property of multiplication there is no real difference between the approaches.

8.1.2 Scales and Scaling

The map scalebar in Figure 8.1.1 shows not only a graphical scale that can be used to directly measure real-world distances from the (scaled) map representation, but also indicates a ratio: 1:31,680. What does this scale mean? Does it have units?

Map scales are like most other dimensionless proportions, as we discussed in Chapter 4. The beauty of many dimensionless proportions is that we can use *any* units we want in them provided that both parts of the scale ratio or proportion have the same units. So for the map scale, we can say that 1 inch on the map is equal to 31,680 inches in the world that the map represents[6]. If we measure out a path on the map that is x inch long, the distance we would cover walking along that path in the real world is $x \times 31680$ inches. That's not a very easy distance to envision because there aren't any very familiar benchmarks near that quantity of inches, but we could convert the latter to feet or miles to make it simpler, and then we can re-express the scale in as a dimensional ratio: $31680 \text{ in.} \times 1 \text{ ft.}/12 \text{ in.} = 2640 \text{ ft.}$. We can go one step further still: $2640 \text{ ft.} \times 1 \text{ mi.}/5280 \text{ ft.} = 0.5 \text{ mi.}$. That works out pretty nicely, and it often does so by design! So we can restate the map scale as 1 inch = 0.5 miles, or equivalently 2

[6] For better or worse, the US still persists with using inches, feet, and miles as conventional measures of distance in official maps and documents. Even though most scientists adopted the metric system long ago, we retain imperial units here to recognize the persistence of legacy units in our maps.

Recall that benchmarking is a process of conceptualizing the size of a quantity by comparing it with a known reference quantity.

1 mile = 5280 feet

inches per mile. This means the same thing as the scale ratio 1:31,680 but is more specific because we have already chosen the units that we wish to measure with. Note that we cannot say that the map scale is 2:1 or 1:0.5 because in converting the second number from inches to miles we've made the scale statement unit-specific.

Maps aren't the only thing that we encounter that are scaled representations of reality. When we are learning about microscopic properties of molecules or cells, we often look at diagrams or physical models of things that are too small to see. When looking through a microscope, we perceive a much enlarged version of the object of study. In each case, we are seeing representations of reality scaled to size that is easier for us to grasp. Importantly, we are also (usually) seeing things scaled *isometrically*, meaning that all dimensions are enlarged or shrunk by the same constant factor. We'll see in later chapters some interesting problems associated with scaling that is not isometric.

8.1.3 *Example: the area of roads in a county (Problem 3.3)*

One reasonable sub-problem to address in the issue of deer-car collisions is how widespread are roads in the area of interest? As with several of the other teaser problems in this book, no specific county is cited, so as a first approximation I'll just estimate numbers for my own home county: Story County, Iowa. According to Wikipedia, Story County has an area of 574 mi^2. The extent of roads in Story County or elsewhere in the US is something that could be readily assessed with a GIS system, and that would certainly be among the most accurate and efficient ways to obtain this value for specific counties. However, for the sake of a first approximation let's try something easier. Literature about road systems in the US suggests that there is about 1.2 miles of road per square mile of land, on average[7]. Clearly this is an underestimate in urban areas, and perhaps an overestimate in remote, rural areas. In the not-so-remote gridded farmscape of central Iowa (Figure 8.3), a road density closer to 2 mi./mi.^{2} is perhaps more appropriate. By this estimate, my county would have approximately $2 \times 574 = 1148$ miles of roads.

[7] This number comes from: Transportation Research Board and National Research Council, 2005. *Assessing and Managing the Ecological Impacts of Paved Roads*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11535.

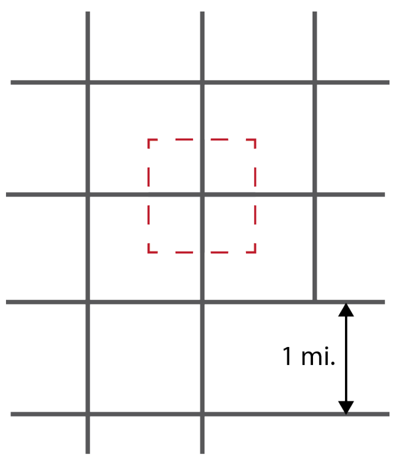


Figure 8.3: In rural farm country, roads are often arranged in an almost-regular grid with spacing of 1 mile. In these settings, we can estimate the "road density" by imagining a square-mile box centered on a road intersection. Within the box shown above, there are 2 miles of road, suggesting a road density of 2 miles per square mile.

Road density is informative, but it only gets us partway to the notion of area. What we need to know now, given that we have road length, is the average width of a road. Let's assume this is 20 feet. To be a meaningful area, we either need to decide to convert length to feet or width to miles[8]. Since the county area was estimated in miles, it would be wise for comparison purposes to convert road width into miles as well. Thus, the average road is about $20/5280 = 0.003788$ miles wide. Multiplying that road width (in miles) by the

[8] This is a seemingly-trivial but still significant decision. In Schoenfeld's framework for problem-solving, making this kind of decision deliberately with the broader goals and practical issues in mind is an example of **control**.

total road length (in miles) yields about 4.35 square miles. That's about $4.35/574 \times 100\% = 0.76\%$ of the county area!

8.2 Geometric Idealization or Approximation

Some problems require spatial measurements or computations that are complex, time-consuming, or difficult to visualize. In some of these cases, a rough estimate may be adequate for the type of solutions we seek; in other cases, we may wish to establish quick ballpark estimates before we dig too deeply into the complex computations, much like we just finished doing in the previous section. For these spatial problems, it can be helpful to idealize the spatial information we have in terms of simple geometric figures that we know something about. For example, suppose we wish to estimate roughly how much an 40cm-long snake might weigh, and we have no prior information or experience upon which to base this estimate. If we are able to estimate its diameter, we may make some progress by idealizing the snake as a long cylinder. Consulting the table below, we find that a cylinder's volume is expressed as:

$$V = \pi r^2 h \tag{8.2}$$

where r is half of the diameter and h is the length of the snake. If the largest r is around 1cm and h is 40cm, a first guess for the total volume is about 126 cm^3. Now if this radius corresponds to the largest part of the snake, this volume might be an upper bound. Since the snake's body tapers a bit, perhaps a mean radius is better – say 0.7cm. Now the volume is 61.6 cm^3.

Next, given that many non-avian animals have densities close to that of water[9], we can estimate the mass of the snake using density and volume. To see how we should do that, we might use a strategy described in an earlier chapter: write out the problem with just dimensions. We're looking for how much it weighs, but really what we want is a mass. If we list the dimensions of the variables we have and want, they look like this:

$$\text{mass:}[M] \qquad \text{volume:}[L^3] \qquad \text{density:}[M\ L^{-3}] \tag{8.3}$$

we see that the mass we are looking to solve for appears in the density term. The volume term appears as a negative 3rd power in the density term and a positive 3rd power in the volume itself, so when multiplied together, volume (V) and density (ρ) should yield dimensions of mass: $m = \rho V$. If we remind ourselves that the definition of density is indeed $\rho = m/V$, this makes sense as a simple algebraic modification of that definition. Thus, our estimated mass for the

[9] in so-called cgs units, the density of water is 1.0 g/cm^3.

snake would be:

$$61.6\text{cm}^3 \times 1.0\text{g}/\text{cm}^3 = 61.6\text{g}. \tag{8.4}$$

Is this close enough? Perhaps, but that depends on the nature of the problem: why do we wish to know how much the snake weighs, and what will we do with that information?

8.2.1 Example: herbicide purchase (Problem 3.2)

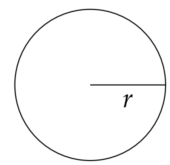
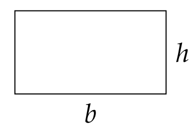
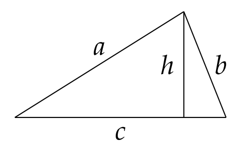
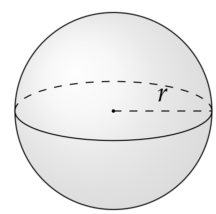
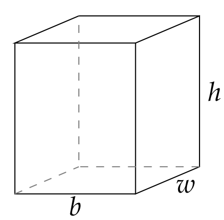
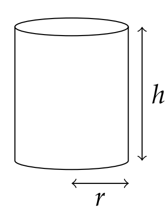
We can use a similar approach to get a start with the herbicide volume needed to eliminate woody shrubs from our city greenspace. A reasonable assumption is that, when cut with a saw or lopper, the exposed stem cross-section of a woody plant is approximately circular. If the goal is to cover the stumps completely with a coating of herbicide, each stump will have a volume (from Equation 8.2) equal to its cross-sectional area πr^2 times the thickness h of the herbicide coating applied to the stump.

8.3 Measuring polygon area

Not all spatial bodies of interest to us are easily measured using the simple idealized shapes reviewed above. Alternatively, geometric idealizations introduce too much error for certain applications. Nonideal shapes can, however, be approximated by irregular polygons in some settings. Here, we describe a method for computing the area of an arbitray two-dimensional polygon using a clever trick that is frequently built into CAD, GIS, and other geospatial software packages. The primary requirement we must meet to use this method is to have coordinate pairs for each vertex in the polygon in a Cartesian (i.e., a plane with two orthogonal, linear axes; a.k.a. an x-y plane) coordinate system. This algorithm may be implemented easily enough by hand, but finding the area of a more complex shape is better left up to a computer. Let's have a look at how this algorithm works.

Table 8.1: Geometric relationships for common shapes.

PROPERTIES OF SIMPLE GEOMETRIC FORMS

shape	property
circle circumference: $2\pi r$ area: πr^2	
rectangle perimeter: $2b + 2h$ area: bh	
triangle perimeter: $a + b + c$ area: $\frac{1}{2}ch$	
sphere surface area: $4\pi r^2$ volume: $\frac{4}{3}\pi r^3$	
rectangular prism surface area: $2bw + 2bh + 2hw$ volume: bwh	
cylinder surface area: $2\pi rh$ volume: $\pi r^2 h$	

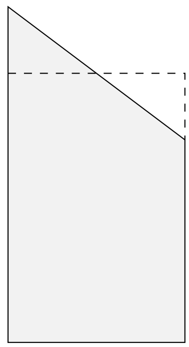


Figure 8.4: A trapezoid (in gray) with dashed line indicating the rectangle with the same area.

First, recall that a trapezoid is a four-sided shape in which only two sides are parallel, as in Figure 8.4. At first glance, it might seem that the area of the trapezoid would be challenging to estimate, but when we realize that it should be the same as the area of a rectangle that's as tall as the average "height" of two vertical sizes of the trapezoid, we may see some hope[10]. We know that the area of a rectangle is just its height times its width: $A_{\text{rect}} = hw$. For a trapezoid whose vertical sides have heights h_l and h_r for left side height and right side height, respectively, we can restate the formula for area in terms of the average of those heights:

$$A_{\text{trap}} = \frac{1}{2}(h_l + h_r)w \tag{8.5}$$

Now suppose that instead of defining heights and widths in terms of h and w, we have the vertices of our trapezoid in Cartesian coordinates[11]. Each point at a vertex (corner) of the trapezoid therefore has an x, y coordinate, where x refers to the horizontal coordinate direction and y is vertical. In this system, note that the width of the trapezoid will be defined by the difference in two x coordinates. So if we take the upper left corner in Figure 8.4 to have an x coordinate of x_1 and the upper right corner to be at x_2, the width of the trapezoid is $x_2 - x_1$. Similarly, if our trapezoid height extends to zero[12] in the y coordinate direction, the "average" height of our trapezoid can be re-written by substituting y_1 and y_2 for h_l and h_r. If we make these changes to the area formula above, any given trapezoid in our coordinate system has an area:

$$A_{\text{trap}} = \frac{1}{2}(y_1 + y_2)(x_2 - x_1) \tag{8.6}$$

Now consider the polygon depicted in Figure 8.5. Each pair of adjacent vertices can be viewed as the upper corners of a trapezoid. If we apply the formula above to each pair of adjacent vertices and add the areas together, what do we get?

$$A = \frac{1}{2}[(x_2 - x_1)(y_2 + y_1) + (x_3 - x_2)(y_3 + y_2) + ... \\ ... + (x_1 - x_n)(y_1 + y_n)] \tag{8.7}$$

Here, n is the number of vertices and the ellipsis "..." means that we've left out some number of terms in the equation, though in this case we've only got five vertices so we've only left out two terms. Notice that if we systematically label the vertices in our polygon in a clockwise manner, about half of the trapezoids will have negative areas and half will have positive areas, though the negative areas will be somewhat smaller. This is because as we come around the bottom

[10] In Figure 8.4 you can imagine snipping the tope of the trapezoid on the dashed line, flipping it over and filling the void in the upper right. Alternatively, we could imagine splitting the trapezoid into a shorter rectangle and a full-width triangle and compute the area as the sum of those two areas. With a bit of algebra, we find that the result is the same.

[11] Cartesian just means that we are specifying the location of points in a two-dimensional coordinate system where the coordinate directions are perpendicular to one another. We will see below that the UTM coordinate system for maps is Cartesian, whereas latitude and longitude are not.

[12] in fact we could show that it needn't go all the way to zero provided that all of the y values in the trapezoid collection are positive all are negative, but this is left as an exercise.

of the polygon as we march clockwise from one pair of vertices to the next pair, our x coordinates are becoming smaller as we go from right to left. This is good! The result is that the width value (and consequently the area) computes as negative, and as a result this lower trapezoid is subtracted from the total area of the polygon, which is exactly what we want! If we order the vertices counterclockwise we'd get the opposite result, but the resulting (negative) area would still be correct.

As mentioned earlier, this algorithm is readily implemented on a computer, using either spreadsheet software like Excel, or computational/statistical software like R. Likewise, this method is built into other software tools that natural resource students and professionals use, including most GIS packages.

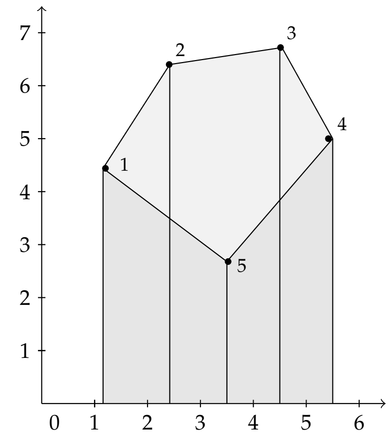


Figure 8.5: An arbitrary polygon. Using the trapezoidal algorithm, areas for individual trapezoids are computed one-by-one in clockwise order. In the lower part of the polygon (where trapezoid fill is darker gray), the computed areas are negative, trimming the unwanted trapezoid area from below the polygon.

8.3.1 Example: open-water waterfowl habitat (Problem 3.1)

One important variable that influences waterfowl abundance is the presence of different types of habitat. Most waterfowl feed extensively in open water, so the area of open wetland is a key habitat variable. In Section 8.3 we identified the trapezoidal algorithm as a tool for estimating the area of arbitrary shapes. We also acknowledged that, while it is possible to do the required computations by hand, automating the algorithm improves computational efficiency by orders of magnitude. The algorithm can be implemented as a formula in a spreadsheet containing the coordinates of the polygon but, as we've discussed, this operation is common enough that it is incorporated in most GIS software. Therefore, comparing watershed areas between the parcels described in Problem 3.1 is a geospatial problem. We nevertheless provide an opportunity in the chapter Exercises below to work with this algorithm directly.

Exercises

1. What is the longest distance across the wetland in figure 8.1?

2. Estimate the dimensionless map scale ratio from the scalebar in figure 8.1.

3. Would the map scale be the same or different if you made an enlarged photocopy of a map?

4. Use geometric approximation to estimate the area of the wetland in figure 8.1.

5. Use the trapezoidal algorithm to make a more precise estimate of the wetland area. The data table below contains UTM coordinates of the wetland perimeter.

	easting, m	northing, m
1	477698	4651450
2	477720	4651488
3	477738	4651524
4	477746	4651549
5	477768	4651554
6	477776	4651574
7	477784	4651607
8	477792	4651630
9	477789	4651652
10	477794	4651668
11	477825	4651670
12	477855	4651675
13	477882	4651671
14	477906	4651678
15	477923	4651701
16	477946	4651714
17	477971	4651721
18	477994	4651723
19	478002	4651737
20	478009	4651761
21	478015	4651779
22	478035	4651794
23	478053	4651792
24	478066	4651792
25	478083	4651794
26	478099	4651794
27	478116	4651789
28	478131	4651780
29	478136	4651757
30	478132	4651736
31	478114	4651714
32	478099	4651694
33	478081	4651678
34	478066	4651661
35	478070	4651633
36	478068	4651604
37	478065	4651587
38	478065	4651562
39	478065	4651539
40	478061	4651516
41	478053	4651500
42	478040	4651491
43	478033	4651472
44	478023	4651449
45	478010	4651425
46	477986	4651406
47	477953	4651394
48	477923	4651379
49	477905	4651371
50	477882	4651364
51	477857	4651356
52	477837	4651359
53	477814	4651368
54	477797	4651363
55	477781	4651368
56	477761	4651368
57	477740	4651368
58	477716	4651376
59	477705	4651391
60	477698	4651412
61	477690	4651432
62	477698	4651450

9
Triangles

9.1 Measuring with Triangles

It is fair to ask why we should bother learning about triangles, since their relevance to ecology and natural resources isn't immediately apparent. Indeed, natural materials tend to approximate more tabular or rounded shapes, and true triangles are harder to find in nature by comparison. But the real power of triangles is not in where we can see them, but where we can *imagine* them. Believe it or not, imaginary triangles can help us measure properties of a landscape or organism, and that fact is firmly incorporated into many of the tools and technologies that researchers and professionals use. In particular, the ratios between the lengths of triangle sides is one of their key assets. In this chapter, we'll review some of the properties of triangles and see how these properties can be leveraged to measure things we care about.

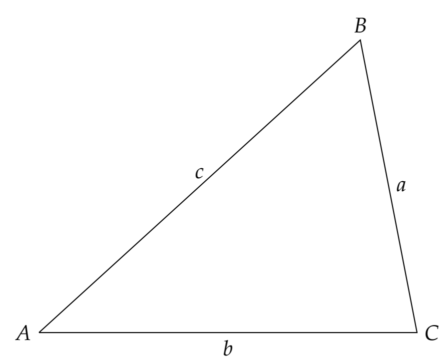


Figure 9.1: A simple triangle. Note that the corners (or vertices) are labeled with capital letters, and that the sides opposite those vertices are labeled with the lower-case version of the same letter. This is partly by convention, and to facilitate some of the techniques we'll review below.

9.2 Trigonometry primer

Trigonometry is the study of triangles, specifically the relationships between the lengths of their sides and the angles between them. At first glance, that might not seem like it is very relevant to the natural sciences, but a few examples might convince you otherwise:

- Determining the distance "as the crow flies" between two geographic points is often most easily done with the help of triangles.
- Measuring the height of a tree or a mountain can employ triangles.
- Telemetry often uses triangulation to determine the geographic location of collared animals.

If you've had a trigonometry class, you might associate the discipline with manipulating equations with $\sec 2\theta$ and $\cot(1 + \pi/2)$. Outside of math class, did you ever find yourself needing to find the secant of an angle? Not likely. But it isn't too uncommon to encounter the likes of sine and cosine, which are often written sin and cos, respectively. That's because they are really, really useful[1]. And it turns out, nearly all the other trig functions you may have learned about are *readily defined using sines and cosines*! For example, the tangent of an angle can be defined as the ratio of the angle's sine to its cosine, but it is so useful that we should recognize it as well.

[1] At this point you may just need to take my word for it, but I hope that you'll appreciate this fact by the time you finish this chapter.

As far as I'm concerned, on the off chance that you ever need to manipulate an equation containing the hyperbolic cosine (cosh) of something, you can look it up or type it into an internet tool like Wolfram | Alpha. If this is your first experience with trigonometry, have no fear, we'll take it slow!

Wolfram | Alpha is a web-based tool developed by mathematician and entrepreneur Stephen Wolfram. It is based on the same underlying computational engine as the math software Mathematica, but can be used (for free with slightly-limited functionality) from any web browser. In addition to performing computation and algebraic simplification, it can attempt to comprehend simple written questions and can retrieve data from a few established databases, concerning for example weather, fincance, and sports.

Similarity

Similarity is a concept that may not boast enough sophistication to warrant its inclusion in a trigonometry class. However, it is an intuitively easy idea to grasp and its utility can be great. And fortunately for modern scientists, the formal application of similarity allows us to design tools for measuring things efficiently in the field.

The principle of geometric similarity[2] is straight-forward. If we have a given shape with known side lengths and/or known angles formed between the sides, we can say that another shape is similar if it has the same number of sides and a relationship between those sides and angles that is the same as our reference shape. The two shapes can still be similar even if they are not the same size or orientation. If any combination of translation (moving the shape), rotation, reflection (a mirror image), or isometric[3] scaling could allow you to overlay one shape on the other to find them to be identical, the shapes are similar.

[2] In some fields this concept is endowed with a more sophisticated sounding name: geometric simili*tude*.

[3] The word **isometric** in this context means that any change in one spatial dimension of a shape (e.g., length) is matched by a proportional change in all other dimensions.

For triangles, the qualifying criteria for similarity are simple, since there are only three sides and one internal angle at each of the three vertices. For shorthand, when two triangles have one angle that is identical, we'll call that A. When two triangles have all three angles equal, we'll refer to that is AAA. Likewise, if the side-length of one side of two triangles is equal, we'll describe that with S. With these definitions, we'll make the following claims, as yet unproven, about the criteria for determining similarity:

TRIANGLE SIMILARITY
Two triangles are similar if any one of the following can be established:

- **AAA**. The angles of one triangle are equal to the angles of the second.
- **SSS**. The side-lengths of one triangle are equal to those of the second. Side lengths may be scaled by a constant C if that constant is the same for each side.
- **SAS**. Two side-lengths and one angle of one triangle are equal to those of the second. Side lengths may be scaled by a constant C if that constant is the same for each side.

When you've reached the end of this chapter, you should be able to show how each of these criteria for similarity could be derived from one of the others. That is left as an exercise for you to work on, one that can build your intuition for using triangle properties for problem solving.

9.2.1 *The right triangle and sohcahtoa*

The mnemonic SOHCAHTOA is one of a small number of things that most trig students remember years after taking the class. Indeed this is really helpful way to remember the algorithms relating the basic trig functions to ratios between the sides of a **right triangle**[4]. But it reveals nothing about the ways that triangles can be employed for practical purposes. So before we deal with these functions, let's revisit what a right triangle is, where we might encounter one, and a few terms and rules regarding these beasts. Figure 9.2 shows a nice, well-behaved one.

[4] As you may know, a right triangle is defined as a triangle with one right, or $90°$ angle.

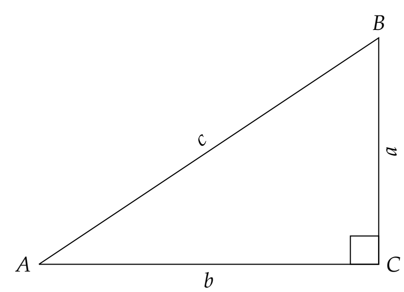


Figure 9.2: A right triangle.

Notice that each **vertex** (a corner point) joins two of the three sides and doesn't touch the side that is opposite it. For reasons that might be apparent in a moment, we choose names for the sides and vertices that imply a relationship between a vertex and the side opposite it. So for example, side a is opposite vertex A (meaning the vertex A is not one of the endpoints of side a). It probably satisfies your intuition that the size of the angle at a vertex might have some simple relationships to the length of its opposite side – at least there is a more intuitive relationship there than between vertex A and one of the other two sides. Imagine keeping vertices A and C stationary, but allowing the angle at A to grow. It is plain to see that if the angle $\angle A$ increases, the vertex B must move up and the length of side a increases accordingly. This thought experiment produces similar results for the other opposing side/angle pairs as well, and we'll use it to our advantage shortly in dealing with triangulation.

One of the most fundamental properties of all triangles is that the three vertex angles sum to 180° ($\angle A + \angle B + \angle C = 180°$). For the special case of a right triangle, the right angle by definition is 90°, so the other two angles must be smaller than 90°. This seems obvious, but it has an important consequence: the longest side of a triangle is the one that is opposite the largest angle. Therefore, since the right angle is the largest angle in a right triangle, the longest side (which we call the **hypotenuse**) is across from the right angle (Figure 9.3).

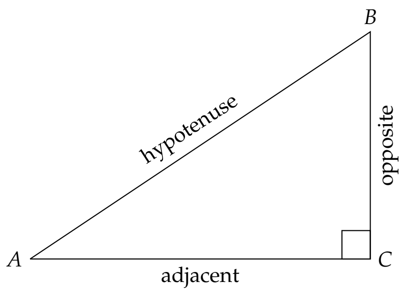


Figure 9.3: A right triangle.

In addition to the rule that angles must sum to 180°, one of the most powerful properties of right triangles is their adherence to the **Pythagorean theorem**:

$$a^2 + b^2 = c^2. \tag{9.1}$$

This is always true provided that c is the hypotenuse of a right triangle. It turns out that there is a simple modification that can be made to this if we are dealing with any arbitrary triangle. But before we go much farther, consider an example setting where a right triangle can be a useful aid to measurement.

9.2.2 *Example: overland distances*

If you are visiting GPS waypoints stored as UTM coordinates[5], the distance on-the-ground between two points might not be obvious from the coordinate sets. For example, what is the distance between (452632,4660214) and (452991,4660580), the marked locations of two observed dickcissel nests? Once we recognize these ordered pairs as the geographical equivalent of (x,y) pairs, it is pretty easy to see that the second nest is $452991 - 452632 = 359$ m east and $4660580 - 4660214 = 366$ m north of the first. But neither a dickcissel nor an ornithologist would likely go from one nest to another by first going 366 m north and then 359 m east. Both would be more likely to go in an approximately straight line. Since east and north are perpendicular, we can construct a triangle like Figure 9.4, with a 359 m long easting side and a 366 m long northing side to represent the coordinate distances. The as-the-crow-flies distance is the hypotenuse of the triangle since it is opposite the right angle. Therefore we can use the pythagorean theorem to find that distance, which we can call d:

$$d^2 = (\text{easting})^2 + (\text{northing})^2 \tag{9.2}$$

$$d = \sqrt{(\text{easting})^2 + (\text{northing})^2} \tag{9.3}$$

$$d = \sqrt{(359\ m)^2 + (366\ m)^2} = 513 \text{ m} \tag{9.4}$$

[5] Universal Transverse Mercator, or UTM, refers to a projected geographic coordinate system wherein locations are given coordinates (meters easting, meters northing) according to their distance in meters east and north of a predefined datum. The benefit of UTM coordinates compared with latitude and longitude is that it is an orthogonal coordinate system analogous to the Cartesian x-y coordinate system we sometimes use for abstraction in math.

Knowing that the second nest is about 513 m away from the first is great. But if you were to give instructions to a field assistant to walk 513 m from the first nest to find the second nest, that alone is insufficient information to get to the correct place. Which direction does she have to go? You could, of course, simply have her walk north 366 m and then east 359 m, but that wouldn't be terribly efficient. What is missing is obviously the direction. If she's carrying a compass, you could give her a bearing or compass direction to follow, but what is that bearing, and do we have enough information to determine it?

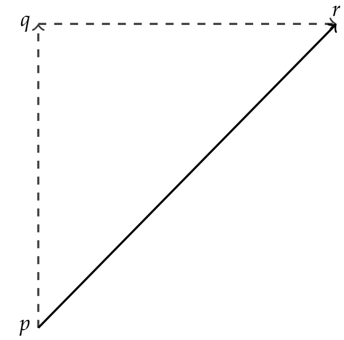


Figure 9.4: Dickcissel nest distances. North is toward the top of the page.

9.2.3 *Angles and azimuths*

At this point, we need to draw more important distinctions regarding coordinate systems and conventions. When we need to be more accurate than simply saying "northeast", compass bearings are often given as angles in degrees. Some people prefer to use **quadrant** bearings, where directions are given with respect to deviations from north or south. For example, due northeast might be expressed as "north 45 east", or equivalently N45°E. That can be interpreted as 45° east of due north. Similarly, southeast could be S45°E and southwest is S45°W. This can sometimes be easier to understand in conversation, but bearings expressed in **azimuth** are less prone to mis-interpretation. Azimuth is the compass direction in degrees clockwise from north, increasing continuously from 0° to 360°. In this system, north is both 0° and 360°, east is 90°, south is 180° and west is 270° (Figure 9.5).

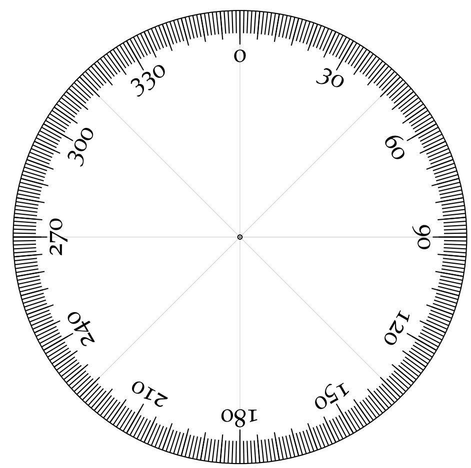


Figure 9.5: The azimuth coordinate system in a compass. Graphic modified from D. Orescanin.

In the world of mathematics, angles are usually measured counterclockwise from the x-axis[6]. Not only does this mean that the starting point (0°) is in a different place, but it increases in a different direction. We will occasionally employ this convention, since it is so prevalent in quantitative topics unrelated to geography. But for the current problem, we'll stick with azimuth.

[6] In math and physics, angles are also frequently measured in **radians** rather than degrees. Note also that many computer programs that are able to do trigonometric computations assume that your argument (or desired result) will be in radians. Radians make alot of sense for geometry because, by definition a radian is the angle traversed when you traverse a length along the perimeter of the circle *that is equal to the radius of the circle*. But for practical use in the field, degrees are easier to work with. So here's a quick rule of thumb in case you need to convert: once around a circle is 360° and 2π radians.

Returning to the problem of finding the dickcissel nest, how can we decide what bearing to give to the field assistant? Since the easting and northing distances are similar, we can be fairly confident that it will be near 45° (NE), but probably not exactly that. But how do you determine the unknown angles in a triangle when you only know one angle (the right angle) and all of the side lengths? Aha! sine and cosine to the rescue!!

9.3 *Angles, circles and sines*

Before we proceed with finding the azimuth, let's more formally define a few trigonometric quantities. We'll do this initially in a mathematics framework, using an x-y coordinate system with angles increasing counterclockwise from the x-axis. Figure 9.6 shows what we might call the "unit circle", a circle centered on the point (0,0) (also called the "origin") with a radius of 1 unit. If we choose any point on the circle, call it p, it lies a distance of 1 unit from the origin. But its coordinates are not immediately obvious. As with the case of the dickcissel nests, however, we can break the path from the origin to p into a component in the x direction and a component in the y direction. Connecting each of those component paths with the direct path (the radius line), we end up with a right triangle $\triangle$OAH, as illustrated in yellow in Figure 9.6.

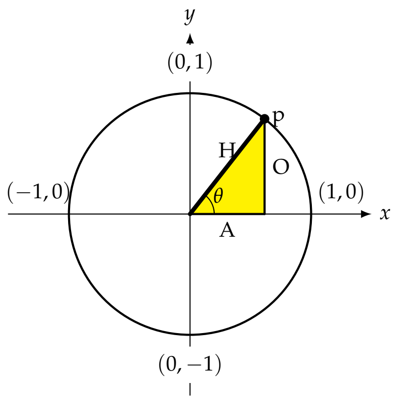


Figure 9.6: Right triangle inscribed inside the unit circle.

Since we have a right triangle, we could use the Pythagorean theorem again to find an unknown side length, provided that we know two of the sides. But in this case, we do not know two sides. Instead we know the angle θ between the x-axis and the line connecting the origin with point p. We also know that this line, call it H for hypotenuse, has length 1 by definition. The other two sides, O and A for opposite and adjacent, are unknown but can be found from the fundamental **trigonometric functions**.

TRIGONOMETRIC RATIOS
The **sine** of an angle is the ratio of the length of the opposite side (O) to the length of the hypotenuse side (H).
The **cosine** of an angle is the ratio of the length of the adjacent side (A) to the length of the hypotenuse side (H).
The **tangent** of an angle is the ratio of the length of the opposite side (O) to the length of the adjacent side (A).

The tangent is identical to the ratio of the sine to cosine of an angle, which you can see is equivalent to the above definition if you cancel the hypotenuse terms in the ratio of ratios. These are loaded definitions, so let's take a few moments to ponder what we've just seen.

- The trig functions are *functions* in the formal sense: they convert an input (angle) to a unique output (ratio of side lengths).
- The opposite and adjacent sides are defined relative to the angle that is the *argument* of the trig function.
- The output of each trig function is a dimensionless quantity that represents the ratio of two side lengths.
- If we know one angle (other than the right angle) and one side length, we can find the two remaining side lengths using the trigonometric functions.

In equations, we don't spell the full name of these functions, but instead use sin, cos, and tan as shorthand. With this shorthand and the definitions above, we can construct a few simple equations that can help us find the unknown side lengths O and A in Figure 9.6:

$$\sin\theta = \frac{O}{H} \tag{9.5}$$

$$\cos\theta = \frac{A}{H} \tag{9.6}$$

Since we know θ and H on the unit circle, we can rearrange these equations to solve for the unknowns. Multiplying both sides of each equation by H, we get:

$$H\sin\theta = O \tag{9.7}$$

$$H\cos\theta = A \tag{9.8}$$

Now the reason we have done this in the unit circle is that $H = 1$, so we essentially end with the definitions $\sin\theta = O$ and $\cos\theta = A$. An important thing to realize, then, is that we can *scale up* to any arbitrary side lengths. Suppose we were interested in not a triangle with hypotenuse 1 unit, but one with hypotenuse 55 meters. Defining $f = 55$ and isometrically scaling all the triangle sides by that factor, we can see for example that:

$$\sin\theta = \frac{fO}{fH}\left[\frac{\mathrm{m}}{\mathrm{m}}\right]. \tag{9.9}$$

Of course the f's simply lengthen H and O by the same common factor, and could be easily canceled. But this illustrates the fact that sine and other trig functions describe dimensionless side-length ratios, and that those ratios can scale proportionally *without changing the sine, cosine, and tangent of the angles*! An obvious implication is that if one vertex of a right triangle has the same cosine and sine as another triangle, the triangles can be shown to be similar.

Actually, we already knew that, since by knowing one angle other than the right angle in a right triangle we can easily find the third. Recall, then, that one of the criteria for identifying similarity in triangles is AAA, or equality of all three angles regardless of side length. For any pair of similar right triangles, the only difference in side lengths is a constant scaling factor f. Thus, any right triangle can be scaled down to a similar one on the unit circle by dividing all side lengths by the length of the hypotenuse, so that $H/f = 1$!

In addition to simplifying triangle scaling, the unit circle also allows us to imagine moving point p along the perimeter and observing how the lengths of A and O change accordingly. In fact, the software Geogebra is perfectly suited for doing this, and I highly recommend playing around with it to boost your intuition. The key thing to notice is that the denominator of the side ratios defined as the sine and cosine is the hypotenuse, or 1 on the unit circle. So the lengths of the opposite (sine) and adjacent (cosine) sides are the output of those respective functions. For angles between 0° and 90°, both functions range from 0 to 1. If you allow x and y to take on negative values as point p goes down or to the left of the origin, you'll see that both functions remain between -1 and 1, inclusive.

But tangent is a different story. Recall that one definition for the tangent of an angle is O/A. You can see that for a small angle θ, O is quite small and A is pretty close to 1, so the ratio of the two will be nearly 0. O and A are equal and $\tan\theta = 1$ when θ is 45°, and as θ approaches 90°, A approaches 0 and O approaches infinity, so $\tan\theta$ approaches infinity as well. What do you suppose happens as θ gets larger than 90°?

9.3.1 *Example: tree clinometry*

One of the most common ways to measure the height of a tree is with a clinometer. This is a small handheld device with a sighting lens and crosshair and one of a variety of different mechanisms for measuring the inclination angle (either positive or negative) of the sight-line from horizontal. Figure 9.7 illustrates the hypothetical triangles constructed with vertices at the observer's eye, and the base and crown of the target tree. Obtaining the height distribution of merchantable timber in our forest parcel in Problem 3.5 could include a set of representative height measurements with a clinometer.

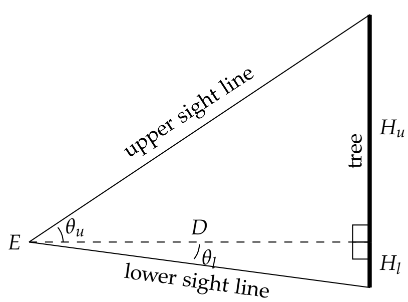


Figure 9.7: Clinometry measurement of tree height.

In measuring the height of a tree, two readings are often taken with the clinometer: one to the crown θ_u and one to the base or stump θ_l. The observation point E is a pre-determined distance D from the tree itself. From this information, what is the height of the tree?

We identify H as the quantity of interest, and observe that $H = H_u + H_l$. As a first step, we must therefore determine H_u and H_l. We assume the geometry of the problem allows us to construct two imaginary right triangles as illustrated and that our clinometer gives us angles in degrees from the horizontal. Since we know the horizontal distance to the tree D and have measured the angles to the top (θ_u) and bottom (θ_l) of the tree, we know the adjacent side length (D) and an angle for both triangles. The target unknowns are the opposite sides of each triangle, and from *sohcahtoa* we know that we can use tan to find the opposite sides when we know the adjacent sides.

Thus:

$$\tan\theta_u = \frac{H_u}{D} \tag{9.10}$$

$$D\tan\theta_u = H_u \tag{9.11}$$

and

$$\tan\theta_l = \frac{H_l}{D} \tag{9.12}$$

$$D\tan\theta_l = H_l \tag{9.13}$$

and since $H = H_u + H_l$,

$$H = D\tan\theta_u + D\tan\theta_l \tag{9.14}$$

$$H = D(\tan\theta_u + \tan\theta_l). \tag{9.15}$$

Thus, we can use an elementary trigonometric function (tan) and a bit of algebra to produce a formula that relates clinometer angle measurements to tree height.

9.4 Arbitrary triangles

While some problems may be approached profitably with imaginary right triangles, others present triangles without right angles. We'll call these **arbitrary** or **general triangles**. Triangulation is a typical task in which ecologists or wildlife managers might encounter arbitrary triangles. When radio-tracking a collared animal, for example, one method for determining the animal's location at a given time is by triangulation from multiple directional antennae. Figure 9.4 is the same as Figure 9.1, except that now we are considering the vertices to be radio transmitters (animals) and receivers (ecologists).

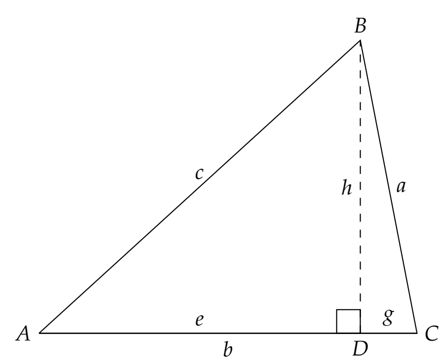


Figure 9.8: Triangulation in telemetry. B is a target radio-collared animal and A and C are directional antennae.

Thus far, we only have two tools that are safe to use with triangles that lack a right angle: the rule that all the interior angles sum to 180° (for which we may use the shorthand ($\Sigma\angle 180$) and the general criteria for and implications of similar triangles. These might be of limited use if our goal is to determine the distance to or location of a collared animal. But if we know the length of one side (say side b in Figure 9.4), and a few angles, we can make some progress.

9.4.1 Law of sines

The law of sines is valid for general triangles, including right triangles. If we are careful to define our sides and vertices as we have (with vertex angle A opposite side a and so on), we can state the law of sines as follows:

LAW OF SINES

$$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}. \tag{9.16}$$

Note that this equation is strange in that there are two equals signs. Don't get worried, this is just a shorthand that allows us to say on one line that each ratio between the sine of an angle and the length of its opposite side is equal to the other corresponding sine/side ratios. If we wrote each equation with a single '=', there would be three of them and it would simply take more space. But in actual implementation you can use any of the implied equalities in Equation 9.16, such as:

$$\frac{\sin A}{a} = \frac{\sin C}{c}. \tag{9.17}$$

The key concept here is that there is a simple and consistent relationship between the sine of each angle and the length of its opposite side, and that this *applies to all triangles, no matter what size or shape*. That makes plenty of sense right? If you imagine taking a triangle formed by three knotted rubber bands and lengthening one side (without changing the length of the other sides), what happens to the angle opposite that stretched side? It grows right? But to accommodate the growth of that angle, the other two angles must get smaller. The law of sines is especially helpful if you know two sides and one angle (SSA) or two angles and one side (AAS).

9.4.2 *Law of cosines*

Another tool that is useful for general triangles is called the "Law of cosines". In many ways, it provides the same information that you can easily find from other tools we have already discussed, so we won't derive it or discuss it in great detail. But mathematician Paul Lockhart makes the case that the law of cosines might be a misleading name, and that the relationship might be better described as a modified version of the Pythagorean theorem that is good for all general triangles. Check it out:

$$c^2 = a^2 + b^2 - 2ab\cos C. \tag{9.18}$$

As you can see, it is identical to the Pythagorean theorem except that there is a correction factor $2ab\cos C$ that is subtracted to account for deviations from a right triangle. As with the Pythagorean theorem, the law of cosines gives you the third side length if you know the other two, but you also need to know the angle between the known sides (SSA). As such, it's function overlaps that of the law of sines.

9.5 Triangle tools: a summary

Trigonometry is a large subdiscipline of mathematics, and can and does fill more than a semester in math classes. Our treatment here has focused on the tools that are most commonly encountered in practical field settings in the natural sciences. Many additional functions, relationships and skills can become important in specific, more technical applications, but most of these can be derived from the basic functions discussed here. These functions and properties are summarized in the table that follows.

Rule name	Relationship	right	general	use for
Angle sum	$\sum \angle = 180°$	X	X	AA known, want A
Pythagorean theorem	$c^2 = a^2 + b^2$	X		SS known, want S
Similarity	$X_1 = CX_2$	X	X	C known, want X (A,S or other)
sine	$\sin\theta = O/H$	X		AS(O or H) known, want S(H or O)
cosine	$\cos\theta = A/H$	X		AS(A or H) known, want S(H or A)
tangent	$\tan\theta = O/A$	X		AS(O or A) known, want S(A or O)
Law of sines	$\frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}$	X	X	AAS or SSA known, want S
Law of cosines	$c^2 = a^2 + b^2 - 2ab\cos C$	X	X	SSA known, want S

9.5.1 Example: shoreline waterfowl habitat (Problem 3.1)

Some dabbling duck species like Mallards seem to prefer very shallow water. This means that small, shallow wetlands can fit the bill, but even the shallow shoreline areas of larger and deeper wetlands may be adequate. Shorelines are also the interface between feeding and nesting areas for many species, and often support diverse flora and fauna across the ecotone.

One way we could estimate the extent of shoreline habitat is to find the length of wetland perimeters, or outlines. If, as described in the last chapter, we have digitized (or obtained existing data for) the outlines of wetlands in our candidate parcels, we should have easting and northing coordinates for these outlines. As with polygon area, most GIS software will automatically compute the perimeter of any shape. However, it is instructive to see how this follows from our earlier discussion of triangle-assisted spatial reasoning.

Recall that when we were traversing between dickcissel nests, we used an implementation of the Pythagorean theorem to find the straight-line distance from the coordinates of both end points. We can write this relationship in an x, y coordinate system as follows:

$$l_{1\to 2} = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]}, \tag{9.19}$$

where $l_{1\to 2}$ is the length of the straight-line distance from point 1 (with coordinates x_1, y_1) and point 2. Notice that the result is always a real, non-negative number because the differences are squared. If we have a series of n points describing the wetland outlines, the sum of all n of the lengths forming a closed polygon approximate the perimeter P of the polygon[7]. We can generalize this as follows:

$$P = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]} + \sqrt{[(x_3 - x_2)^2 + (y_3 - y_2)^2]} + \ldots$$
$$\ldots \sqrt{[(x_1 - x_n)^2 + (y_1 - y_n)^2]}. \quad (9.20)$$

As with the trapezoidal area formula, this can be implemented by hand, in a spreadsheet, or with GIS software.

[7] How good is this sort of approximation of the perimeter? This is a simple question with a not-so-simple answer. For our purposes here, the more points we have, the better – particularly if we've got an automated way to do the computations. However, in a more philosophical sense this is the crux of the coastline paradox, first popularized by mathematician Benoit Mandelbrot.

Exercises

1. Use Figure 9.4 and the right triangles formed by dropping the perpendicular h, to derive the law of sines from the trig functions you already know.

2. Explain how you could find the telemetered location of B if you know the locations of A and C and their internal angles.

3. What happens if you try to apply the law of sines to a right triangle?

4. What happens if you use the law of cosines on a right triangle? Assume angle C is the right angle.

5. Triangles can be used to measure distant objects, even if we can't get to them. This can be used to estimate the height of an object (e.g., a mountain peak or tree) where we cannot access the base, and therefore cannot measure the full horizontal distance that separates us from the object we wish to measure. The general premise of the remote method is illustrated in figure 9.9. If we can use a clinometer (a device for determining the angle between horizontal and a sight-line to an object of interest) to determine the sight-line angles to the object of interest p from two different places $b1$ and $b2$, and we know the distance between those places l, we can use trigonometry and algebra to determine the desired height H.

   Devise a strategy for measuring H from the information gathered at $b1$ and $b2$. Derive and justify a formula that can be used for this task.

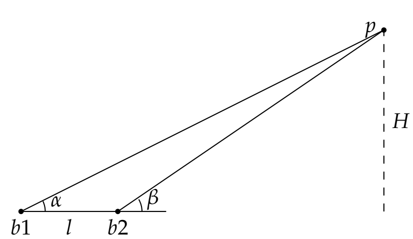


Figure 9.9: Triangle abstraction of the mountain problem.

Part IV

ALGEBRAIC REASONING

10

Generalizing Relationships

We have encountered many instances in this book where solving a problem numerically required numbers that we didn't have. We often don't know alll the numbers needed to solve real problems. In some cases, the simplest way to overcome this issue is to estimate or guess a value. However, in many other cases, the value of an important quantity isn't constant in space or time – our lack of knowledge is not a reflection of uncertainty in measurement. Instead, there is a systematic variation in the real value of a quantity and we need to allow for those changes. Under these circumstances we need to treat these quantities as **variables** that have unknown numerical values. Perhaps we have some idea of how large or small the numberical values can get[1], but within these limits, the variable can take on any value.

[1] For example, if we've used ballpark estimates and deliberately chosen high-end estimates of some of the parameters, this could provide us with an approximate upper limit on the value of the variable of interest.

In this new world of uncertainty, we have the tools of algebra at our disposal. At least in parts of the problem-solving process, this can be disorienting as we have to carry symbols rather than reducible values through any operations that we find necessary. However as we'll see shortly, performing symbolic manipulations as a means to solving problems can lead to versatile and reusable solutions. What we have done prior to now can be called **specializing**, where we seek particular numerical values in every calculation when possible. The alternative, which you'll recognize as a stepping-stone for algebra, is to **generalize**.

HEURISTIC: Express variable quantities needed for solving a problem as symbolic variables and manipulate them according to the rules of algebra to yield **general** relationships.

WRITING ALGEBRAIC RELATIONSHIPS can seem to be hocus pocus at first. However, the mystique fades a bit when we remind ourselves that mathematical relationships are little more than formal logical statements. By carefully assembling what we know about quantities of interest, striving to sustain generality, and following a few tips, we can begin to use algebraic reasoning as a powerful tool for creativity and sense-making.

Writing algebraic expressions

- Identify the relevant variables and constants
- Introduce descriptive notation for each quantity
- List what is already known about each variable, using expressions with symbolic notation when possible
- Look for ways to set expressions equal to one another based on what you know; are there two ways to define the same quantity using the variables of interest?
- Guess or infer unknown relationships
- Write and simplify equated expressions as a symbolic equation
- Check for dimensional or unit consistency

When we express and manipulate equations with symbolic variables, we are doing algebra. When we state systematic relationships between symbolic variables, we're using functions. Functions can describe derived, hypothetical, or observed relationships, depending upon how we arrive at them.

10.1 *Families of Functions*

In the natural and environmental sciences, a few families of functions can be used to describe relationships between key variables of interest. We will explore the most prevalent of these kinds of functions, examining their algebraic composition and the characteristics of their graphs. In the chapters that follow, we wil see how functions can be used to describe relationships between measured variables and how they can be used to devise mathematical models.

10.1.1 *Linear functions*

The simplest relationship between two variables – let's call them x and y – is perhaps something like $y = x$. This relationship is indeed a linear relationship, stating only that y is equal to x without any modification, or that any change in the variable x results in an identical change in y. In reality, we will rarely encounter any relationships like this that are worth describing in an equation. Instead we may often find that the variables of interest are related through a constant of proportionality, call it m (might as well stick with the notation we may have seen elsewhere!). In this case, $y = mx$ is still a linear relationship, but now for any change in x, we expect a change in y that is m times as large as the change in x. That is what this *function* does for us: it converts any proposed value of x into a corresponding y according to the definition of the function. Indeed, the definition of a function in mathematics is an operation that takes a value of an **explanatory** or **independent variable** as input and produces a value for a unique **response** or **dependent variable** as output.

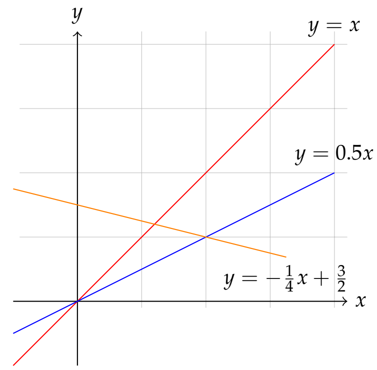


Figure 10.1: Various linear functions.

Here's an example: an elephant's tusks grow continuously with age, beginning a bit less than a year after birth. Although there is likely some variability within the population, this relationship allows biologists to estimate elephant age. Thus, a mathematical description of this relationship can be written as a linear equation:

$$l = ra \tag{10.1}$$

where we are calling tusk length l and age a. Notice that this way of writing the relationship implies that we are treating a as an independent variable (that is, we can think of it as sort of a **cause**) and l is the dependent variable (an **effect** that depends on the cause). Depending on the circumstances, these roles could be switched. Indeed, it is easier to measure tusk length than age for a given elephant, so we might wish to use measurements of tusks to help determine the age distribution in a wild elephant population.

Also important to remember is that when we are doing science instead of just math, the variables usually have units and dimensions, which we discussed previously. If we expect our equation to be meaningful, the dimensions on the left- and right-hand sides of the equation need to be consistent (i.e. equal). So in the elephant tusk example $r = l/a$ is a growth rate, must have dimensions of length per time, or $[L\ T^{-1}]$ (see how we get that? would that be the same if we swapped our independent and dependent variables?). If we have been measuring length in inches and age in years, our value for r should be in inches per year.

Great! But as we said above, we might wish instead to know the age as a function of tusk length. So we need to rearrange things a bit. Let's now define m as the number of years of age per inch of tusk length, which is just the reciprocal of r. In other words, $m = 1/r$, which also means $r = 1/m$. Since we've just taken a reciprocal here, the dimensions of m are just the reciprocal of the dimensions of r, $[T\ L^{-1}]$.

Now we can re-state our new linear relationship as $l = a/m$, or $a = ml$. In this form, we have the dependent variable (age) on the left-hand side of the equation and the independent variable (tusk length) on the right hand side, as is convention. But at this point, what is implied about an elephant's age if it's tusk length is zero ($l = 0$)? Regardless of the value of m, plugging zero into this equation yields $a = 0$. Of course, as mentioned above, adult tusks do not begin to develop until several months after birth. So our equation is probably not very good at representing reality (particularly for young elephants), and is therefore not yet useful. But suppose we were to change what we mean by "age" on the left-hand side. It makes sense that what we're measuring is growth from the age when the tusks first appeared, so let's call that age a_0, which is close to 0.5 years. So the elephant age that we wish to determine is more than we would have predicted before by the an amount corresponding to the age when the tusks first appeared, a_0. So our new equation, modified to account for this correction, reads:

$$a = a_0 + ml \tag{10.2}$$

In the abstract but precise terms of mathematics, we say that a is a linear function of l with a **slope** of m and offset (or **y-intercept**) of a_0. Although it seems obvious in the context of this example, the offset a_0 must have dimensions of time for this equation to be meaningful. Note that the value of the growth rate, or slope m, can be determined algebraically by solving the linear equation above for m:

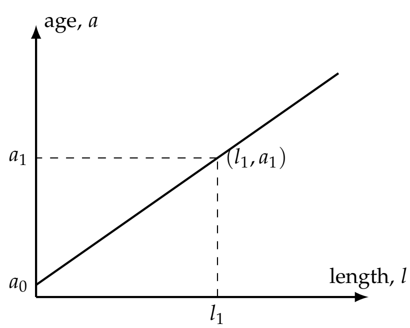


Figure 10.2: Plot of tusk length l as a function of age a.

$$m = \frac{a - a_0}{l}, \tag{10.3}$$

which we might recognize as the "rise" of the function, $(a - a_0)$ divided by the "run" l.

In some cases it is unnecessary, but in others we may need to specify something about the **domain** (a set of upper and lower constraints on the values of the independent variable) over which a proposed relationship is valid. For example, it doesn't make sense for an elephant to have a negative age any more than it does to have a negative tusk length. There is probably an upper limit to tusk length as well, though it is hard to be confident what that might be. To be complete but conservative, we may specify that the domain of the function as $0 < l < 160$ inches. The **range** for our linear function is the spread of minimum to maximum values of a corresponding to the minimum and maximum values in the domain. Note that this last comment applies to linear functions (though sometimes the signs are reversed), but for some non-linear functions of interest, maximum and minimum values in the range may not correspond to maximum and minimum values bounding the domain. We'll see examples of this later.

Functional relationships that are approximately linear are very common in the sciences. Indeed, a routine procedure in the analysis of multivariate data is **linear regression**, wherein the coefficients (slope and intercept) of a linear function that best fits the data are sought. Linear functions – or nonlinear ones for that matter, as we'll see below – can also be postulated hypothetically in the construction of mathematical models.

10.1.2 Example: fire suppression costs (Problem 3.5)

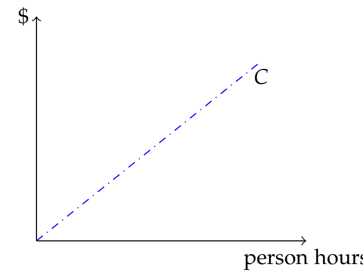


Figure 10.3: Schematic illustration of a hypothetical linear relationship between the cost of fire suppression and the number of person hours of suppression effort.

The issue of how much suppression effort to use is at the heart of this problem, so it's clear that suppression effort should be considered a variable. As is routine with variable quantities, we should assign a symbol to the variable and decide, at least for now, what units it will be quantified with. A single symbol is preferable (though subscripts are permissible if necessary) to prevent any ambiguity. Therefore, let's choose the symbol E for effort, and provisionally assign the units of person-hours. A person hour has dimensions of $[1\ T]$. Like acre-feet and other similar units, person hours is compound unit that we should understand as the number of hours worked per person multiplied by the number of people. For example, if two people work 8 hours each, that effort represents 16 person hours.

Now we need to deal with the other variable that is implied in this part of the problem: cost. First, suppose we define C as the symbol we'll use, and US dollars as the unit of cost. Relating the cost of suppression to the effort requires some way of assigning a cost per unit of effort. Recalling our choice of units, this cost-per-unit-effort will have units of dollars per person hour, which sure sounds like an hourly wage. In fact, that's exactly what it is! So let's call it w. We can now state the algebraic equivalent of the sentence "suppression cost equals the number of person hours of effort times the hourly wage". In symbols:

$$C = wE \tag{10.4}$$

This equation is illustrated in Figure 10.3 as a straight line increasing from left to right. The slope of the line, analogous to m in our abstract concept of the prototypical linear function, is w, and the y-intercept is zero. This latter observation simply articulates the (hopefully obvious) notion that the cost of zero person hours of labor should be \$0.

10.1.3 Polynomial functions

A polynomial function is one in which the dependent variable also depends on the independent variable raised to an exponent. Polynomials are among those functions that can have multiple ups and downs in the dependent variable over the domain of the function. To refresh your memory, let's write an abstract polynomial equation for starters:

$$y = b + mx + lx^2 + kx^3 + ... \tag{10.5}$$

Here we have a function in which the quantities added together on the right-hand side have dependence on increasing powers of x, and the way we've written it we imply that the equation could go on indefinitely, incorporating ever-growing powers of x as we go. One way to describe a polynomial is by its *order*, which is nothing more than the integer values of the exponents of x included. If we take away the "..." from the equation above and just stop the equation after kx^3, this would be a third-order polynomial, since 3 is the highest exponent of the independent variable x. Sometimes you will see the term *cubic* for third-order polynomials, while the term *quadratic* is used for second-order polynomials. To be complete, we can even pretend that the first term on the right-hand side b is really bx^0, representing a "zeroth order" term.

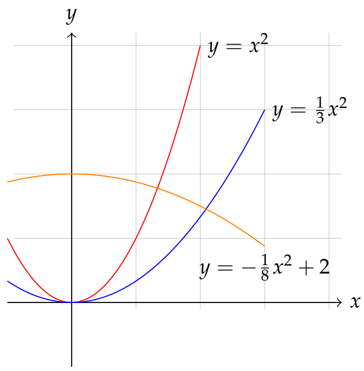


Figure 10.4: Polynomial functions.

Suppose that we got rid of the l and k terms in the equation above, which we could simply do by saying that $l = k = 0$. What's left is just the linear function we had above, and we see now that the linear function is really a special case of a polynomial function, a "first-order polynomial". Likewise, if $b = m = k = 0$, we're left with just $y = lx^2$, a second order or quadratic polynomial. It is still a second-order polynomial if m and b are nonzero. So you see that writing the equation as we did above allows us to imagine a polynomial of any order that we choose, and keeping or discarding any terms we wish by adjusting the lettered coefficients.

So how do these polynomial functions differ from linear functions? Take as an example the formula for the surface area of a sphere – perhaps representing a raindrop: $A_s = 4\pi r^2$. A simple linear function as described above has a single independent variable and the values of the dependent variable depend only on the first power of the independent variable and a constant of proportionality. We can write the surface area equation as $A = (4\pi r)r$, and now it looks like we only have the first power of r. Great, but now our constant of proportionality contains r, so it is not a constant at all but a variable itself. So **nonlinear functions** are those that cannot be written as a relationship between the dependent variable and the first power of the independent variable times a *constant* constant of proportionality.

Before we waste too much more time talking about polynomials, I need to be clear on one thing: When we encounter polynomials in most undergraduate mathematics classes, we are only considering functions where the powers of the independent variable are *whole integers*. With this in mind, it is worth thinking about whether they are really useful for us. What relationships depend on integer powers of the independent variable? One area where these polynomials are useful in natural science is spatial measurement. you probably remember that the areas of squares and circles each depend on the second power of a characteristic length (side or radius). Likewise, volumes of spheres and cubes depend on the 3rd power (coincidence?). While we may never encounter perfect spheres and cubes in the natural world, we may find occasion to idealize the size and shape of something (like a sand grain, egg or raindrop as a sphere, a tree root or snake as a cylinder, etc.) in a simple model so that we can better understand something about it.

Likewise, some physical phenomena can be described with equations that depend on a whole number power (often 2) of time or position. In more complicated problems in the real world, it can also be advantageous to approximate an unruly function using a so-called *series expansion* of the function, which often amounts to a polynomial.

These examples notwithstanding, true polynomial functions do not arise as commonly in the natural sciences as linear and some other non-linear functions do[2]. An important exception, which we'll grapple with quite a bit later this term, is the so-called logistic or density-dependent growth function. In ecology, this function describes the theoretical growth of populations constrained by limited space or resources. We can write the basic relationship as:

[2] But as we'll see below, there are certainly relationships in the natural sciences where the relationships between variables are best described with functions that have *non-integer* exponents.

$$G = rN - \frac{r}{K}N^2, \tag{10.6}$$

where the dependent variable G is the population growth rate $[1\ T^{-1}]$, the independent variable N is the number of individuals, r is a growth constant and K is the carrying capacity. This function is quadratic because N^2 is the highest power term.

10.1.4 *Power functions*

In addition to linear and polynomial functions, it is relatively common to encounter at least four other classes of functions in the natural sciences. **Power functions** arise commonly in ecology and geography, especially in scaling properties of organisms and habitats in space. Power functions may include any function in which the independent variable is raised to an arbitrary exponent, of the form:

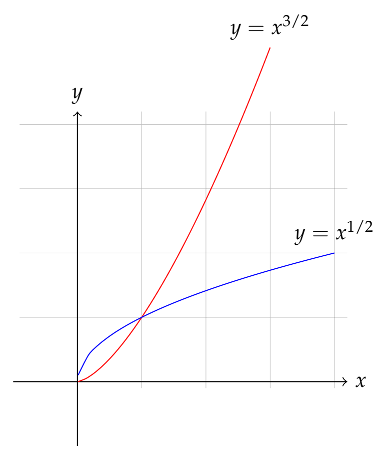


Figure 10.5: Examples of power functions.

$$y = ax^b \tag{10.7}$$

The power function differs from a polynomial in that the exponent on the independent variable is not constrained to be an integer. Figure 10.5 compares the appearance of power functions with exponents greater than and smaller than 1.

In the biological subdiscipline of island biogeography, a relationship between island area A_i and species diversity S has often been described with a power law:

$$S = cA_i^z \tag{10.8}$$

where c is a fitting parameter and z is an exponent that is usually less than 1.

Another example of a power function appears in the description of what hydrologists call a stream's "hydraulic geometry", which describes how the width, depth and average velocity of a river change in time and space[3]. Channel width w, for example, typically increases downstream in a way that can be described as $w = aQ^b$, where Q, the water discharge, is our independent variable and a and b are empirical constants. Channel depth and velocity are described with analogous relationships.

[3] One of the scientists who developed and popularized this concept was Luna Leopold (1915-2006), the second son of Aldo Leopold.

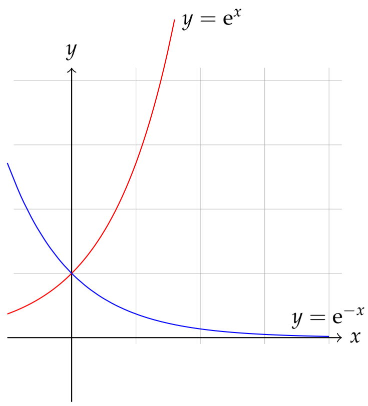


Figure 10.6: Examples of simple exponential functions.

10.1.5 *Exponential functions*

Exponential functions are different from power functions in that the independent variable appears as part of the exponent, rather than the base. A generic exponential function might look something like this:

$$y = a^{bx}. \tag{10.9}$$

The base may often be e, which is an important (but irrational like π) number close to 2.718, but needn't be. Exponential functions describe ever-increasing or ever-decreasing change, and appear in contexts like the decay of radioactive substances or unrestrained growth of populations. The radioactive decay equation might look a bit like this:

$$N = N_0 e^{-\lambda t} \tag{10.10}$$

A similar form describes the extinction (attenuation) of sunlight with depth in a water column or forest canopy according to the Beer-Lambert law:

$$I = I_0 e^{-kd} \tag{10.11}$$

where d is the independent variable. In addition to natural growth and decay phenomena, exponential functions appear extensively in economic analysis.

A somewhat more complicated form of exponential function is sometimes used to describe growth of individuals (fish, trees, etc.) over time. The Von Bertalanffy growth function (VBGF) can be written:

$$L_t = L_\infty[1 - e^{-K(t-t_0)}]. \tag{10.12}$$

In general, when the exponential argument is a negative number, these functions describe decay or asymptotic approach toward a limiting value. However, when the argument is positive, exponential functions describe explosive growth.

10.1.6 Logarithmic functions

Closely related to exponential functions are **logarithmic functions**. The natural logarithm, sometimes written ln, is the inverse function of e, meaning that $\ln(e^x) = x$. The base-10 logarithm, written $\log_{10}$ or simply log, behaves in a similar way but for exponential functions with base 10. So $\log_{10}(10^x) = x$. Both types of logs, and logarithms with any other base, are functions that increase rapidly for low values of the independent variable, but increase ever slower thereafter. We will find logarithms especially useful in transforming data that we suspect might be a power or exponential function, and must therefore have a basic command of the algebraic rules that apply to them. Outside of transformations and inverting exponentials, however, we won't encounter logarithms extensively.

Exercises

1. Given the description of species-area relationships given in Section 10.1.4 and the notion that the exponent z in Equation 10.8 is less than 1, describe what this means conceptually. How does the species diversity change with island area, and how does an increment of area change affect small islands differently than larger islands?

2. Using only symbolic variables and constants, write an expression that defines that time necessary for 95% of a radioactive isotope to decay. *Hint: interpret this to mean that we seek an expression for t when* $N/N_0 = 0.05$.

3. Review the description of Problem 3.2. Write a hypothetical, but well-justified, algebraic equation relating the volume of herbicide needed to eliminate woody shrubs, and the basal area per unit land area of those shrubs. Consider all quantities to be variables, so use symbols rather than numbers for this.

4. Review the description of Problem 3.1. Using reasonable geometric idealizations (not computer algorithms), can you write a simple algebraic equation that relates the length of a wetland's perimeter habitat to the wetland's area?

11
Relationships Between Variables

In the previous chapter, our discussion of variables and functions largely assumed that relationships were known or developed independent of any measurement or data. However, functional relationships between variables can also be derived from data. Here, we explore two concepts that help us understand the strength and nature of systematic relationships between variables.

11.1 *Correlation*

In common parlance, the word correlation suggests that two events or observations are linked with one another. In the analysis of data, the definition is much the same, but we can even be more specific about the manner in which events or observations are linked. The most straight-forward measure of correlation is the *linear correlation coefficient*, which is usually written r (and is, indeed, related to the r^2 that we cite in assessing the fit of a regression equation). The value of r may range from -1 to 1, and the closer it is to the ends of this range (i.e., $\mid r \mid \rightarrow 1$), the stronger the correlation. We may say that two variables are positively correlated if r is close to +1, and negatively correlated if r is close to -1. Poorly correlated or uncorrelated variables will have r closer to 0.

In the margin are two plots comparing life-history and reproductive traits of various mammals. In the first one, Figure 11.1, the arrangement of points in a band from lower left to upper right on the graph is relatively strong, corresponding to a relatively high r of 0.73. In contrast, the correlation between litter number per year and litter size in Figure 11.2 is (surprisingly?) weak, producing more of a shotgun pattern and r a modest 0.36.

In the abstract, the mathematical formula for the correlation between two variables, x and y, can be written:

$$r = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)\left(\frac{y_i - \bar{y}}{\sigma_y}\right), \tag{11.1}$$

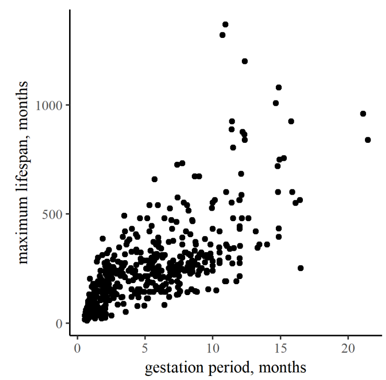


Figure 11.1: Correlation between the maximum lifespan and gestation period of various mammals, $r = 0.73$.

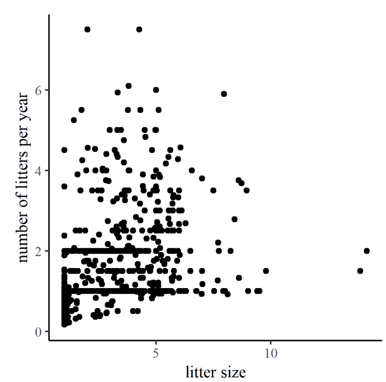


Figure 11.2: Correlation between the number of litters per year and the litter size of various mammals, $r = 0.36$.

where the subscript i corresponds to the ith observation, the overbar indicates mean values, and σ_x and σ_y are standard deviations. The specifics of this formula are not of great interest to us. The important thing to understand is that when positive changes in one variable are clearly linked with positive changes in a second variable, this indicates a good, positive correlation, $r > 0$. The same is true if negative changes in one variable correspond to negative changes in the other. However, positive changes in one variable corresponding to negative changes in another indicate negative correlation, $r < 0$. It is also important to note that this is a good measure of correlation only for linear relationships, and even if two variables are closely interdependent, if their functional dependence is not linear, the r value will not be particularly helpful.

Nevertheless, correlation can still help us identify key relationships when we first encounter a dataset. Consider the changes in weather variables measured at a meteorological station as a function of time. Weather data can be very overwhelming due to the number of variables and the sheer volume of data. One handy way to isolate some of the strongest interdependencies among variables of interest is to look for correlations. A correlation matrix plot is essentially a grid of plots where each variable is plotted against all the other variables in a square array of panels. Relationships with strong positive or negative correlations immediately jump out, suggesting which relatinships we might wish to investigate further. For example, let's look at a month-long weather dataset downloaded from www.wunderground.com.

There is alot of information in these plots, so let's look at them piece by piece. Notice that the panels on the diagonal from upper left to lower right would be a variable plotted against itself ($r = 1$), and they are therefore replaced by a density distribution for each variable. Also notice that since the upper right half would be a mirror image of the lower left, there is just a number in each of those panels rather than a plot. In any case, here we have just selected four variables of potential interest, and you can immediately see that there is a strong positive correlation between mean temperature and mean dew point, with $r = 0.962$. The strength of the correlations from these plots (the six in the lower left) is indicated by the correlation coefficient in the (mirrored) corresponding panel in the upper right. There are also relatively strong negative correlations between temperature and pressure, and dew point and pressure. In contrast, we see weak correlations between humidity and temperature and humidity and pressure, as indicated by the low r values.

The important thing to remember from all of this is what the correlation coefficient can tell us: a high, positive correlation between

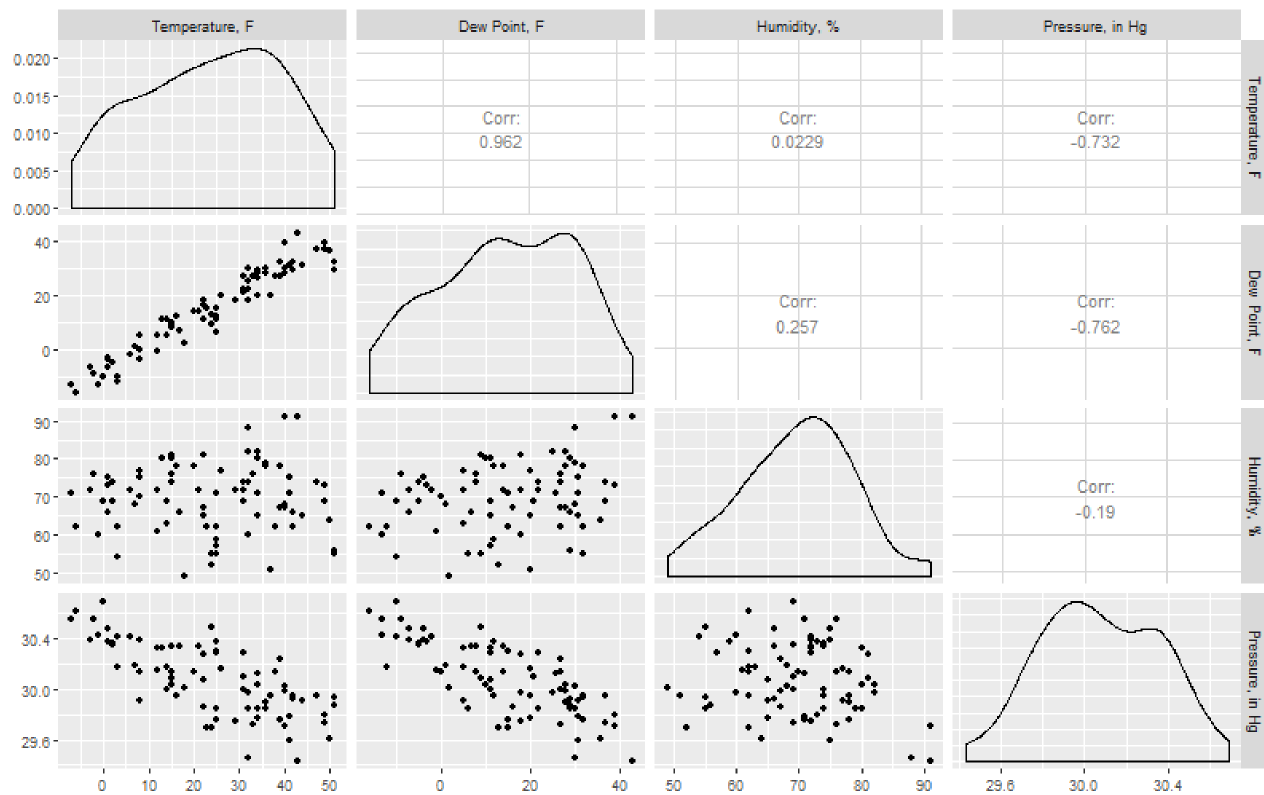


Figure 11.3: Correlation plots for weather data from Ames, IA, April 2014.

two variables indicates that when one goes up, so does the other. A high negative correlation indicates that as one goes up, the other goes down. Low correlation coefficients indicate that a consistent linear relationship cannot be established. If correlation is established, however, this analysis doesn't yet provide details about the functional relationships present.

11.2 *Regression*

Regression is the process of fitting a mathematical function to a set of data points using some criterion for judging "goodness-of-fit". The resulting "best-fit" function may then be used to predict unknown values, to forecast future values, or to evaluate the dependence of one variable upon another. Goodness-of-fit can be determined by one of many statistical techniques that determine how well a function describes the variations in the data used to generate it. The most common criterion for goodness-of-fit is called "least-squares", so you might sometimes see the whole process called **least-squares regression**. Least squares means what it sounds like, sort of. When a function (let's write it $y = f(x)$) is tested for goodness-of-fit, the difference between the y-values predicted by the test function, each of which we can call $\bar{y}_i$, at a given x_i, and the y_i-values in the data set are found, squared, and added together for the entire dataset. The best-fit line is then the one for which the sum of the squares of the residuals are minimized (least). This is very commonly done for linear equations, but we can use the same techniques for nonlinear equations as well.

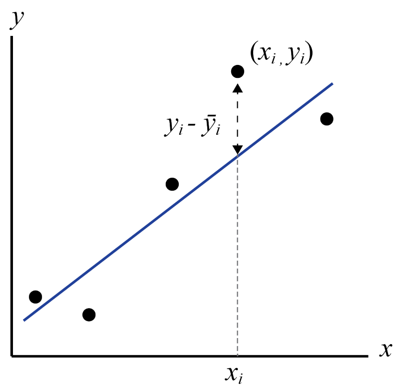


Figure 11.4: Schematic representation of the quantities involved in finding best-fit functions by least-squares regression.

Some data sets that we may encounter just don't appear to have linear trends though. In these cases, we can try transforming one or both variables[1] or we can attempt to perform nonlinear regression. As with many of the statistical and spatial methods discussed in this book, the heavy lifting for most of these options can – and probably should – be done with computer software. However, we should still be aware of what is happening

[1] Some common data transformations include logarithmic, exponential, and reciprocal. In these transformations, a modified variable is created by performing the selected operation on the original variable values.

cumulative catch	catch/effort
86	2.46
137	1.76
169	1.14
178	0.29

11.2.1 *Example: brook trout electrofishing (Problem 3.7)*

Having isolated the age-0 brook trout from each electrofishing traverse and computed the catch per unit effort c_{ue} of that subset, we may now employ the **Leslie method** to estimate the total population of age-0 brook trout in the study reach. In this method, we create a dependent variable $c_{cumul.}$ corresponding to the cumulative number of fish removed in each pass, the "cumulative catch". We then plot and perform a linear regression of the catch per unit effort as a function of cumulative catch, as illustrated in Figure 11.5.

By the Leslie method if we extrapolate the best-fit line to a vertical-axis value $c_{ue} = 0$, the cumulative catch value where that occurs is the estimated total population. This value can be estimated from the graph itself, but the result is better if we solve the for the value directly from the best-fit line. The equation of the best-fit line for this regression is:

$$c_{ue} = -0.0208c_{cumul.} + 4.38. \tag{11.2}$$

Note that the slope of this line (-0.0208), consistent with intuition, is a negative number. The y-intercept 4.38 corresponds to the hypothetical initial catch per unit effort at the very start of the first traverse. Rearranging and solving for $c_{cumul.}$ gives

$$c_{cumul.} = 210.6 - 48.1c_{ue} \tag{11.3}$$

and we find that the estimated total population is 210.

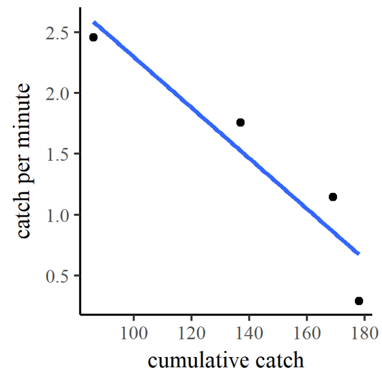


Figure 11.5: Catch per unit effort as a function of total catch for age-0 brook trout, from Table 11.2.1.

Exercises

1. Discuss in a paragraph the benefits and drawbacks of deciding, prior to any data analysis, what type of function to seek best-fit parameters for.

2. In Section 11.2.1, we skipped several steps in the algebraic manipulation that allowed solution for $c_{cumul.}$. Carry out all the intermediate steps, showing your work completely, and determine whether the solution cited above is acceptable.

3. Find a dataset that interests you within a public ecological or natural resource data repository[2], identify variables within a dataset that may be related, and perform a regression to see the nature of that relationship.

[2] For example, browse the Global Registry of Biodiversity Repositories.

Part V

MODELING

12 Modeling

12.1 What is a model?

A model is a representation of reality that allows us to understand something better. There are many types of models, including *conceptual*, *mathematical*, and *physical* models. A physical model is a physical object or set of objects intended to represent something else that is too large, small, complex or otherwise inaccessible for direct investigation. A conceptual model is a collection of hypothesized relationships between different objects or variables, and is usually described in narrative. From an early age, we learn how to construct both physical and conceptual models. Children create conceptual models to help them understand cause and effect relationships that lead to either desirable or unwanted outcomes ('if I jump down one or two steps, it's fun, but if I jump down three or more steps it hurts my legs: jumping farther hurts more'). When my gradeschool son builds a spaceship from Legos, he is creating a physical model of a spaceship he has seen in a movie or book. These are not particularly sophisticated models, but they are nevertheless ways of representing some aspect of reality (or imagined reality).

As with Legos, mathematical models can serve mostly a desire for creative play. Like Lego models, it is perfectly possible to create a mathematical model that represents reality poorly, and is therefore not very useful. Perhaps we claim to have created a model of a car, but if we've only stacked rectangular bricks together and failed to add wheels, it is not a particularly good or useful model of a car. Thus, model construction and use should be done with the broader problem context in mind. The means should justify the desired ends.

Heuristic: Mathematical models are only as useful as the conceptual models on which they are based.

In this book, we are interested in mathematical and conceptual models and the connections between them. Ultimately, our goal isn't necessarily to become mathematical modelers, but rather to be able to construct, use, and understand models that can assist with problem-solving. Indeed, many mathematical models originate from a desire to quantify the relationships in a conceptual model devised to address a problem. Several possible approaches to quantification lead to a handful of varieties of mathematical models. We'll focus our discussion on three distinct but related types of mathematical models that differ in their origins and implementation. The first two are grounded in theory, while the third often arises from statistical data analysis.

- **Analytical models** are usually developed from theory based on fundamental physical, chemical or biological principles. A hypothesis that a tree's height should scale with it's trunk diameter raised to the 2/3 power in order to retain structural integrity is such a model. These models are often the most general and abstract, and can sometimes be solved with paper and pencil. However, they can become hopelessly complex and un-solvable when one tries to incorporate realistic details and context. The idealizations necessary to make an analytical model solvable can also sometimes limit its utility.
- **Numerical models** may be created and motivated in the same manner as analytical models, but employ techniques for mathematical *approximation* that permit relaxation of analytical idealizations and introduction of detail without making the equations too difficult to solve. Numerical models can be solved by hand for very small systems, but are more appropriately implemented in computer programs.
- **Empirical models** may have analytical or numerical components, but contain parameters that must be quantified by experiment or systematic observation. Data must be incorporated and usually analyzed statistically in order to define parameter values. In some cases, regression is used to constrain the functional relationships between variables or to identify the value of coefficients. Thus, a fully empirical model is data-driven or data-calibrated.

We have already seen or worked with a few examples of models. The Logistic population growth model that we discussed briefly in Section 10.1.3 is a theoretical model that can be implemented either in numerical or analytical form. Even that model, however, has empirical components, since it's use in practical problem-solving requires some observational constraints on r and K. When we solved for total brook trout population in Section 11.2.1, we employed an empirical model known as the Leslie method, which is based on a conceptual model of the change in catch probability under declining population.

12.1.1 Example: The Universal Soil Loss Equation (USLE)

The widely-used Universal Soil Loss Equation (USLE) is an example of an empirical model. The master equation for USLE is:

$$A = RKLSCP \tag{12.1}$$

where A is the soil loss (usually in tons/acre/year), R is a rainfall-erosivity factor, K is a soil erodibility factor, L and S are the slope length and angle factors, C is a ground-cover factor and P is a parameter that accounts for soil conservation practices or structures.

The factors in USLE are quantities whose values cannot be measured directly. Instead, the numerical values are each derived from a combination of carefully-designed field experiments where all but one factor is held constant. The factor values are then derived from measured differences in soil loss.

The great value of the USLE and it's kin is that it is sufficiently easy to use that farmers with little formal training in math or computing can easily get satisfactory results. Most factor values can either be looked up in tables or measured on the ground or from maps.

The ease of use comes at a cost, however. Because factor values are derived from experiments, they are strictly valid only within the range of conditions considered within the experiments. In other words, if applied in settings where – for example – rainfall intensity is twice as large as the largest observed in experiments, the reliability of results is uncertain. Fully empirical models can therefore sometimes be unreliable in conditions outside the range of the conditions under which factor values were determined.

12.1.2 Example: probability of deer-automobile encounters (Problem 3.3)

As we have already seen, simple theoretical models can sometimes be sufficient to explore a range of system behaviors, even when functional relationships are uncertain. These models will inevitable by limited in power by the simplifying assumptions or idealizations used, but when the science or management problem permit a solution with substantial uncertainty, this approach is still warranted.

Let's assume that deer in our county are randomly distributed in space, and that they have no particular reason to either avoid or seek out roads. Call the total area of the county A_c and the proportion of the area occupied by roads f, so that the area of roads $A_r = fA_c$. Let's assume that there are N_0 deer in the county. It follows that – if the deer are randomly distributed – there will be approximately fN_0 deer on the road at any moment. What is that number according to the numbers we produced earlier for Story County, IA? The value of

f was estimated to be approximately 0.0076, so if there are say 1000 deer in the county, we should expect either 7 or 8 of them on the road at any given time. That seems reasonable, but that isn't what we're after. We'd like to know about how likely collisions are between deer and automobiles. So we need to work in something about the number and distance of car trips through the road system, right? This is left as an exercise for the student, as there are many possible ways to approach this.

12.2 *Dealing with higher mathematics*

Many powerful mathematical models have been devised to explore and describe phenomena in nature. Some of the most powerful are those that allow predictions of unobserved or future events or patterns. These can directly inform management decisions provided that managers trust and understand their results. Unfortunately, many of these powerful models employ mathematical concepts and methods that are beyond the typical undergraduate training in math. Does that mean that most people are doomed to never understand or use these models? Absolutely not! There isn't any inherent reason that students need to take calculus, linear algebra, or differential equations courses before they can comprehend the gist of a model constructed with those skills. It certainly helps to have at least a conceptual grasp of some key concepts in calculus, but that doesn't translate to a pre-requisite.

12.2.1 *Example: prairie dog plague (Problem 3.4)*

Since this problem deals with hypothetical future events, it may not be possible to glean the answer directly from past work or from observation. Instead, we can construct a simple model of the prairie dog community with random, probabilistic interactions among well-mixed individuals.

A common way to model disease transmission is with a **compartment model** often called SIR. We consider individuals in a population to be in one of three (or four) states: Susceptible (S), Infected (I), and Removed (R) or Recovered. Individuals move from compartment S to compartment I by disease transmission. Infected individuals in compartment I then either recover and move to compartment R, or are removed from the population by death or isolation. These transfers between compartments are often described with a system of *differential equations*:

$$\frac{dS}{dt} = -\beta SI \tag{12.2}$$
$$\frac{dI}{dt} = \beta SI - \gamma I \tag{12.3}$$
$$\frac{dR}{dt} = \gamma I \tag{12.4}$$

These differential equations are not easily solved in most cases, but we can use them as a basis for a *numerical simulation* of disease dynamics if we are able to estimate the parameters β and γ. A numerical representation of the first equation might look something like this, for example:

$$S_{t+1} = S_t - \beta S_t I_t \tag{12.5}$$
$$I_{t+1} = I_t + \beta S_t I_t - \gamma I_t \tag{12.6}$$
$$R_{t+1} = R_t + \gamma I_t \tag{12.7}$$

This says that in a given time increment, susceptible individuals are moved from the S compartment to the I (infected) compartment at a rate that is proportional to the product of the numbers of individuals in each compartment and the transmission rate constant β. You can see in the first and second equations above that when a number of idividuals infected according to the βSI term in lost from the S compartment (because it is negative), it is gained (positive) in the I compartment. All individuals are accounted for in moving into or out of the I compartment. Similarly, individuals move from the I compartment to the R compartment at a rate governed by the rate constant γ. Selection of these rate constants to a large extent governs the behavior of the model, and thus the predicted fate of the prairie dog colony. But implementing management options informed by positive model outputs is where the biggest challenge arises.

12.3 Power-Law Scaling

CONSIDER THIS SEEMINGLY INNOCUOUS QUESTION: are larger animals heavier than smaller animals?

> YOU: Hmmm, well, yeah I think so?! An adult bear weights more than a snowshoe hare, for instance.

OK, great, but how would we know if this is true more generally? And what exactly do we mean by larger? Does that mean taller? Larger volume? This brings up a few issues that become important when we're talking about real quantities rather than abstract variables. Unambiguously defining quantities can be an important first step in communicating quantitative information. In the next section we'll be specific about what information is required to fully define a quantity. For now let's agree that we're satisfied with relating the mass of an animal to its volume. Do animals that take up more space (i.e., have greater volume) also weigh more? Maybe we can say it another way: is the weight or mass of an animal proportional to its body size? We could write this in symbols:

$$M \propto V? \tag{12.8}$$

The symbol '$\propto$' between M (body mass) and V (volume) means "proportional to". So this isn't an equation yet because we're not sure anything is equal. And of course it's nonsense that an animal's weight is equal to its volume. There must be some other parameter that transforms an animal's volume into a mass. Let's call it c, and try it out in an equation:

$$M = cV \tag{12.9}$$

But what is c? As we said above, we'd prefer to have some meaning for the symbols we throw around in equations. Let's use one of our old algebraic tools for manipulating equations and "solve the equation for c". By that we mean get c onto one side of the equation all by itself. To get there, we just need to divide both sides of the equation by V, yielding:

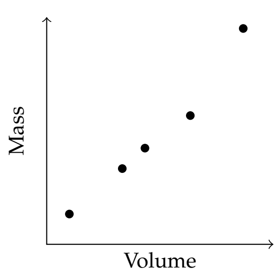


Figure 12.1: Plot of some hypothetical measurements of animal mass and volume.

$$\frac{M}{V} = c \tag{12.10}$$

Now recall that the definition of density is mass per unit volume. That's exactly what we have on the left-hand side of the equation! So our equation now says that c, the parameter we used to transform volume into mass, is the same as density! So for an individual animal, the parameter that relates mass to volume is density. As we have done previously, let's assume that most animals have a density close to that of water so this proportionality parameter c doesn't vary significantly among species. So to the extent that it is correct to say that most animal's body density is close to that of water, we can argue that larger animals do indeed weigh more, in general.

This is probably not a very profound revelation to you[1]. But with only a few more small leaps in logic, we can get somewhere considerably more interesting. For more than a century, biologists have been intrigued by a remarkable relationship between the basal metabolic rate and body mass for animals of a wide range of sizes and shapes. Amazingly, if one assembles a large set of data and plots it on a graph with a logarithmic scale, mice, humans and elephants and most of the rest fall along a straight line! An equation that describes

[1] An entertaining and well-composed article on some not-so-obvious consequences of size differences in animals is *On Being the Right Size*, byt J.B.S. Haldane, published in *Harper's Magazine*, March 1926.

this relationship and the line on the graph looks like this:

$$B = B_0 M^b \tag{12.11}$$

where B is the basal metabolic rate, M is body mass as before, and B_0 and b are constants (we'll see what they mean later!). This equation is yet another power law, and equations with this form pop up surprisingly often in ecology once you start looking. We'll get more into functions and power laws later on. But for now, some important points should be made:

- The argument that there should be a proportionality between body mass and metabolic rate was originally conceived theoretically on the basis that energy given off by an animal to its surroundings might depend mostly on the animal's surface area, while its mass scales with volume.

- Measurements by many researchers over more than a century have been compared against this theoretical prediction, with varying degrees of success. In most cases however, the power-law relationship holds.

- By comparing theoretical predictions with real data, one can discover truly novel and interesting things about physiological similarities or differences between different organisms – insights we might not have ever developed without the quantitative analyses.

  We'll look into this in more detail a bit later.

13
Models of growth and decay

Some of the most well-known applications of quantitative analysis in the life sciences relate to describing changes in processes or ecosystem properties with time. Among the most important examples is population change, where the number of individuals N in a population is expressed as a function of the independent variable t: $N = f(t)$. In this chapter we will explore two types of exponential functions and a polynomial function that form the basis for describing and predicting population change and a lot more.

13.1 *Exponential functions & population models*

An exponential function is one in which the independent variable appears in the exponent, or power, of some other quantity. The equation $y = a^x$ is an example of a simple exponential function if x is the independent variable and y is the dependent variable. In this case, the constant a can be called the *base*, since it is the quantity that is raised to a power. From our high school math classes, we learned about exponential and logarithmic (the inverse of exponential) functions mostly with bases of 10 and e, where e is Euler's number (~ 2.718) and is sometimes written exp(something). But we can have an exponential function with any arbitrary base.

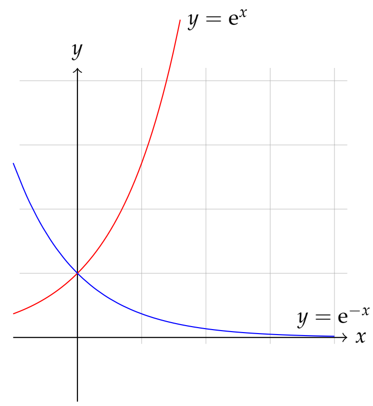


Figure 13.1: The typical ever-changing growth and decay of the exponential function.

Exponential functions arise frequently in economics, physics, and in some contexts in ecology. Imagine, for example, a population of marbled murrelets in a coastal bay in the Pacific Northwest[1]. At some time, suppose their population was 100 individuals. With time, this can change as individuals die or reproduce. If we assume no murrelets emigrate or immigrate (are added to or subtracted from the population), changes in population with time are controlled only by birth and death rates, and we can say the population N after one year is:

$$N_1 = N_0 + B - D \tag{13.1}$$

In this equation, we take N_0 to be a constant, initial population. The

[1] Why murrelets you might ask? As you'll see shortly, it is convenient to begin with "simple" populations, where the causes of population changes estimated from visual surveys are limited.

birth and death rates may scale with the population, such that they can be represented like this:

$$B = b \times N, \qquad D = d \times N \tag{13.2}$$

where b and d are birth and death rates per individual. So, for example, if the birth rate is approximately 0.15 individuals per murrelet per year[2], and death rate is 0.05 individuals per murrelet per year, we can write our equation for population as:

[2] Note that this birth rate is given per individual. Obviously males cannot give birth to offspring, so a better way to express fertility or fecundity is in terms of birth rates per female; however the per individual or *per capita* birth rate is easier to work with.

$$N = N_0 + 0.15N_0 - 0.05N_0 \tag{13.3}$$

If we simplify the right-hand side of this, we have N after one year as a simple function of N_0:

$$N = (1 + 0.15 - 0.05)N_0 \tag{13.4}$$

$$N = 1.1N_0 \tag{13.5}$$

If you plug in 100 for N_0, this gives us an unsurprising result that population is 110 murrelets. This makes sense, since we get $0.15 \times 100 = 15$ births and $0.05 \times 100 = 5$ deaths during that year.

Now if we project into future years (where t is the number of years after our initial measurement of population N_0) with the same relationship, we'll see that after another year of births and deaths, we'll get:

$$N_{t=2} = 1.1(1.1N_0) \tag{13.6}$$

where the quantity in parentheses is the population after one year, now incremented by another series of births and deaths. We can rearrange that equation slightly to yield:

$$N_{t=2} = N_0 \times 1.1^2 \tag{13.7}$$

After another year, we'll get:

$$N_{t=3} = N_0 \times 1.1^3 \tag{13.8}$$

And by now you probably see the pattern. If t is the number of years after an initial population census N_0, our projection of population is:

$$N_t = N_0 \times 1.1^t \tag{13.9}$$

Interpreted as N as a function of t, this is an exponential function with a base of 1.1 and a constant N_0. Note that a very similar function could describe compounding interest on a loan, savings account or credit card balance, if the principal (the amount saved or borrowed) remains unchanged over time.

WE COULD HAVE WRITTEN our equation above a bit differently. Instead of keeping a constant reference to N_0, we could have said that population next year depends only on the population this year and the birth and death rates this year. This alteration would give us:

$$N_{t+1} = N_t + B - D = N_t + N_t(b - d) = N_t(1 + r) \tag{13.10}$$

where $r = b - d$ can be defined as the population's **intrinsic growth rate**. There is no difference in the result of this equation if we apply the same assumptions and constraints as we did in the first version, but this form of the equation is a bit more versatile. It will also become useful to us in a few days. We can call it a *discrete difference* equation.

Before we move on, notice a few things about our population model. First, population is unrestrained. The only factors influencing the growth rate are birth and death rate, and these are considered constants. In reality, these might not be constant as individuals compete for limited resources. Alterations to this model to account for this fact will be introduced next time. Also, notice that the intrinsic growth rate r is positive because we have said that the birth rate is higher than the death rate. It is, of course, possible for the reverse to be true: death rate could be larger than the birth rate, and the resulting r would be negative. As you can see from the above equations, a negative r would result in an exponential decrease in population with time.

When $r = 0$, we may say that the growth rate is zero and births balance deaths. The birth rate that balances death rate is sometimes called "replacement", since it replaces each death with a birth.

13.1.1 More exponentials

One place where exponential functions appear in the natural sciences is in animal physiology, particularly where processes are regulated by temperature. The "surface area" theory for metabolic scaling discussed above suggests that basal metabolic rate scales **allometrically** with the mass of the animal. As we hinted at above, this hypothesis stems from the postulate that metabolic rate scales with the surface area (through which heat can be lost), which is in turn a function of $[L^2]$, where $[L]$ is a characteristic length of the animal. Mass, however, scales with the volume of the animal, which is a function of $[L^3]$. If we combine the two relationships to express metabolic rate as a function of mass, we get the allometric relationship:

$$B \propto M^b \tag{13.11}$$

where B is metabolic rate, M is body mass, and b is the scaling exponent, which is equal to $2/3$ according to the surface area theory. We briefly acknowledged that several studies in the 20th century suggest that the $2/3$-power scaling is not correct, and that a $3/4$-power scaling might be more appropriate. Nevertheless, the general form of the relationship is reasonable. To transform this proportionality into an equation, we could introduce a constant B_0, so that we have

$$B = B_0 M^b \tag{13.12}$$

If we interpret B as the dependent variable and M as the independent variable, this is clearly a *power function* because M is the base. Contrast this type of equation with the population equation above, where the independent variable t was the exponent.

The simple power-law equation for metabolic rate has some simple applications for which it is useful, but it fails to describe many important phenomena that are seen by animal physiologists. One is the fact that metabolic rate is also very sensitive to temperature. A modification to the simple power law was proposed not too long ago in this Science paper. The modification supposes that metabolic rate depends on the kinetics of biochemical reactions on a cellular scale, which are in turn temperature dependent. In chemistry, the temperature dependence of reactions is often expressed as an exponential function of temperature through the Arrhenius relationship:

$$R \propto e^{-\frac{E}{kT}} \tag{13.13}$$

where R is a reaction rate constant and E/k is an energy-related constant for a given reaction, and T is temperature. While this looks a bit ugly, it is an incredibly important relationship for chemistry, physics, and now biology, because it does a surprisingly good job of describing how temperature affects physical and chemical processes.

Let's look for a moment at the general form of this equation by imagining a similar function

$$R = e^{-1/T} \tag{13.14}$$

where we consider temperature T to be the independent variable. As you can see, as temperature increases, the exponent becomes smaller and approaches zero. Since $x^0 = 1$ for all x, this function approaches 1 as temperature increases, but becomes very small for small T. Of course, we cannot compute $1/T$ for $T = 0$, and for that reason the Arrhenius equation is written for T in Kelvin rather than Celsius.

In any case, a much improved relationship for the basal metabolic rate of animals that includes both a dependence on body mass and temperature can be written:

$$B \approx B_0 M^b e^{-\frac{E}{kT}} \tag{13.15}$$

This is a more complex function because it contains two independent variables (mass and temperature), but can be visualized by treating one of them as a constant while the other varies. If we imagine how metabolic rate changes for a single ectothermic organism of a given mass as body temperature changes, it might have a pattern that looks similar to the plot above, but that approaches a value of B_0 with increasing temperature.[3]

[3] If you're interested in more on this topic, revisit this neat article written for the Nature Education Project, and the references therein, or check out this summary of the paper that examined this function.

13.2 *Adding complexity*

Our first population growth model was a simple exponential one. We assumed unrestrained growth with a constant per-capita (per individual) rate parameter $r = b - d$, where b and d are per capita birth and death rates. Our year-to-year prediction of population N with this growth model is

$$N_1 = N_0(1 + r) \tag{13.16}$$

Given an initial population N_0, the population after t years was

$$N = N_0(1 + r)^t \tag{13.17}$$

While we arrived at this result with just some reason and algebra, a more general solution can be found using calculus. We won't worry too much with how this solution is obtained, nor will you be expected to reproduce it, but it is always nice to see how more advanced topics can help us with the problem at hand. So here is a quick summary of how the calculus version works:

If we re-write our first incremental population change equation above

$$N_1 = N_0 + rN_0 \tag{13.18}$$

$$N_1 - N_0 = rN_0 \tag{13.19}$$

Notice that the left-hand side is now just the population *change* over one year. One of the strategies of calculus that allows elegant solution of complex problems is to imagine "smooth" changes, where the increment over which those changes are measured in vanishingly small. While this is obviously an oversimplification of population dynamics (i.e., many animals have discrete breeding seasons so that births are clustered during a relatively small period of time, and no births occur during the remainder of the year), but in many cases we don't need to worry too much about this. We express these vanishingly-small change increments with *derivatives*, where the derivative of N with respect to t can be translated as the instantaneous rate of population change as a function of time, i.e., the population growth rate. With this strategy, the above equation is written:

$$\frac{dN}{dt} = rN \tag{13.20}$$

Applying some second semester calculus, we'd come up with the following solution, which works at all t:

$$N = N_0 e^{rt} \tag{13.21}$$

Compare this equation with the one above, $N = N_0(1+r)^t$, which we developed with discrete differences. Graph both functions and see if they match reasonably well. They should be close, but not exactly the same. The discrete model is, in fact, subtly different, and is often called the *geometric* model for population growth, while the exponential version is the classical *Malthusian* model.

Calculus aside, the above unrestrained population models are useful as a starting point, but they neglect any mechanisms of slowing population growth. In most settings, resource limitation slows or reverses growth rates as population increases. If you're not familiar with the story of St. Matthews Island reindeer, it is an interesting illustration of this effect taken to an extreme.

A fairly simple way to account for resource limitation, and to thereby restrain population growth according to some carrying capacity K, is to include an "interaction" term for our growth rate. Using the same notation as above, an increment of growth in this new population model is:

$$N_1 = (1+r)N_0 - \frac{(1+r)N_0^2}{K} \tag{13.22}$$

This looks a bit clunky, but we can clean it up with a little bit of algebra and by making the same kinds of calculus-oriented modifications that we made above:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) \tag{13.23}$$

As above, the derivative term on the left hand side is the rate of population change as a function of time, or the population growth rate. If we write the equation with G for growth rate on the left-hand side, it looks a bit more manageable:

$$G = rN\left(1 - \frac{N}{K}\right) \tag{13.24}$$

$$G = rN - \frac{r}{K}N^2 \tag{13.25}$$

As you can see, the growth rate is just a second-order polynomial equation. As such, it's graph might be a bit familar to us: it is a downward-opening parabola that crosses the x-axis at $x = 0$ and $x = K$. This is the **logistic** population growth model, perhaps the simplest way of incorporating density dependence and carrying capacity into the description of population changes in a place with finite resources.

Solving this differential equation is not particularly easy, but fortunately for us, smart people have found useful solutions. The most straight-forward solution for N as a function of t is:

$$N = \frac{N_0 K}{N_0 + (K - N_0)e^{-rt}} \tag{13.26}$$

Here is an example of a case where we can defer to the experts who came before us and simply borrow their result for our own use. The fact is, even with the above solution, there is plenty of complexity in the logistic population model since we must define, for any particular scenario, several of the parameters before we can use it to any avail: K, N_0, and r.

13.2.1 Example: minimizing suppression and loss costs (Problem 3.5)

The hypothetical functions we have proposed for the suppression cost C and net value change V_{nc} were simple idealizations and would need to be modified according to better understandings of cost-effort relationships. Nevertheless, our cost-plus-net-value-change function can still allow an instructive optimization. Our function reads:

$$C + V_{nc} = wE + V_0 e^{-kE}, \tag{13.27}$$

where the first term on the right-hand side is the cost of suppression activities, while the second term is the net value change in case of fire. The lowest-cost state is clearly the bottom of the dip in Figure 3.2, but can we identify that point algebraically? If we use a little calculus, we can indeed.

In first-semester calculus, we learn that the maxima and minima of functions can be found by setting the derivative equal to zero. In this case:

$$\frac{d}{dE}(C + V_{nc}) = w - kV_0 e^{-kE} = 0. \tag{13.28}$$

For our purposes here, I won't explain how we arrive at this, but suffice it to say that when we solve the right-hand equality for E, we retrieve the effort corresponding to the minimum total $C + V_{nc}$. We'll follow the algebraic manipulations through here:

$$w - kV_0e^{-kE} = 0 \tag{13.29}$$

$$w = kV_0e^{-kE} \tag{13.30}$$

$$\frac{w}{kV_0} = e^{-kE} \tag{13.31}$$

$$\ln\left(\frac{w}{kV_0}\right) = -kE \tag{13.32}$$

$$E = -\frac{1}{k}\ln\left(\frac{w}{kV_0}\right) \tag{13.33}$$

This result isn't necessarily pretty, but it provides a robust analytical solution that depends only on the coefficients we assigned to the trial functions, and that can be easily modified for different coefficient values.

Exercises

1. Review Section 13.2.1. In the equation for cost plus net value change, there is a constant k. What are it's units?

2. Propose some reasonable values for the constants and coefficients for the fire-suppression problem in Section 13.2.1 and determine the optimal effort and its cost.

3. Review Section 12.1.2 and ensure that you are comfortable with the analysis presented there – or that you have developed and justified your own approach to achieving an analogous solution. Propose and execute a strategy for incorporating car trips through the county road network in order to estimate the probable number of collisions in a given span of time.

4. Review Section 12.2.1. Construct and evaluate a spreadsheet model to solve the numerical approximation of the SIR system of equations.

Index