1

The Basics

1.1 Objectives

This chapter gives a very basic introduction to practical mathematical and arithmetic principles. Some readers who can recall their earlier training in high school and elsewhere may want to skip it or merely skim over the vocabulary. However, many often find that the various other interests in life push out the assorted artifacts of functional expressions, logarithms, and other principles.

Usually what happens is that we vaguely remember the basic ideas without specific properties, in the same way that we might remember that the assigned reading of Steinbeck's *Grapes of Wrath* included poor people traveling West without remembering all of the unfortunate details. To use mathematics effectively in the social sciences, however, it is necessary to have a thorough command over the basic mathematical principles in this chapter.

Why is mathematics important to social scientists? There are two basic reasons, where one is more philosophical than the other. A pragmatic reason is that it simply allows us to communicate with each other in an orderly and systematic way; that is, ideas expressed mathematically can be more carefully defined and more directly communicated than with narrative language, which is more susceptible to vagueness and misinterpretation. The causes of these effects include multiple interpretations of words and phrases, near-synonyms, cultural effects, and even poor writing quality.

The second reason is less obvious, and perhaps more debatable in social science disciplines. Plato said "God ever geometrizes" (by extension, the nineteenth-century French mathematician Carl Jacobi said "God ever arithmetizes"). The meaning is something that humans have appreciated since before the building of the pyramids: *Mathematics is obviously an effective way to describe our world*. What Plato and others noted was that there is no other way to formally organize the phenomena around us. Furthermore, awesome physical forces such as the movements of planets and the workings of atoms behave in ways that are described in rudimentary mathematical notation.

What about social behavior? Such phenomena are equally easy to observe but apparently more difficult to describe in simple mathematical terms. Substantial progress dates back only to the 1870s, starting with economics, and followed closely by psychology. Obviously something makes this more of a challenge. Fortunately, some aspects of human behavior have been found to obey simple mathematical laws: Violence increases in warmer weather, overt competition for hierarchical place increases with group size, increased education reduces support for the death penalty, and so on. These are not immutable, constant forces, rather they reflect underlying phenomena that social scientists have found and subsequently described in simple mathematical form.

1.2 Essential Arithmetic Principles

We often use arithmetic principles on a daily basis without considering that they are based on a formalized set of rules. Even though these rules are elementary, it is worth stating them here.

For starters, it is easy to recall that negative numbers are simply positive numbers multiplied by -1, that fractions represent ratios, and that multiplication can be represented in several ways ($a \times b = (a)(b) = a \cdot b = a * b$). Other rules are more elusive but no less important. For instance, the **order of operations** gives a unique answer for expressions that have multiple arithmetic actions. The order is (1) perform operations on individual values first, (2) evaluate parenthetical operations next, (3) do multiplications and divisions in order from left to right and, finally, (4) do additions and subtractions from left to right. So we would solve the following problem in the specified order:

$$\begin{aligned}
2^3 + 2 \times (2 \times 5 - 4)^2 - 30 &= 8 + 2 \times (2 \times 5 - 4)^2 - 30 \\
&= 8 + 2 \times (10 - 4)^2 - 30 \\
&= 8 + 2 \times (6)^2 - 30 \\
&= 8 + 2 \times 36 - 30 \\
&= 8 + 72 - 30 \\
&= 50.
\end{aligned}$$

In the first step there is only one "atomic" value to worry about, so we take 2 to the third power first. Because there are no more of these, we proceed to evaluating the operations in parentheses using the same rules. Thus $2 \times 5 - 4$ becomes 6 before it is squared. There is one more multiplication to worry about followed by adding and subtracting from left to right. Note that we would have gotten a ***different*** answer if we had not followed these rules. This is important as there can be only one mathematically correct answer to such questions. Also, when parentheses are nested, then the order (as implied above) is to start in the innermost expression and work outward. For instance, $(((2 + 3) \times 4) + 5) = (((5) \times 4) + 5) = ((20) + 5) = 25$.

Zero represents a special number in mathematics. Multiplying by zero produces zero and adding zero to some value leaves it unchanged. Generally the only thing to worry about with zero is that dividing any number by zero ($x/0$ for any x) is ***undefined***. Interestingly, this is true for $x = 0$ as well. The number 1 is another special number in mathematics and the history of mathematics, but it has no associated troublesome characteristic.

Some basic functions and expressions will be used liberally in the text without further explanation. Fractions can be denoted x/y or $\frac{x}{y}$. The **absolute value** of a number is the positive representation of that number. Thus $|x| = x$ if x is positive and $|x|$ is $-x$ if x is negative. The square root of a number is a radical of order two: $\sqrt{x} = \sqrt[2]{x} = x^{\frac{1}{2}}$, and more generally the **principle root** is

$$\sqrt[r]{x} = x^{\frac{1}{r}}$$

for numbers x and r. In this general case x is called the **radican** and r is called the **index**. For example,

$$\sqrt[3]{8} = 8^{\frac{1}{3}} = 2$$

because $2^3 = 8$.

1.3 Notation, Notation, Notation

One of the most daunting tasks for the beginning social scientist is to make sense of the ***language*** of their discipline. This has two general dimensions: (1) the substantive language in terms of theory, field knowledge, and socialized terms; and (2) the ***formal*** means by which these ideas are conveyed. In a great many social science disciplines and subdisciplines the latter is the notation of ***mathematics***. By notation we do not mean the use of specific terms per se (see Section 1.4 for that discussion); instead we mean the broad use of symbology to represent values or levels of phenomena; interrelations among these, and a logical, consistent manipulation of this symbology.

Why would we use mathematics to express ideas about ideas in anthropology, political science, public policy, sociology, psychology, and related disciplines? Precisely because ***mathematics let us exactly convey asserted relationships between quantities of interest***. The key word in that last sentence is ***exactly***: We want some way to be precise in claims about how some social phenomenon affects another social phenomenon. Thus the purchase of mathematical rigor provides a careful and exacting way to analyze and discuss ***the things we actually care about***.

★ **Example 1.1: Explaining Why People Vote**. This is a simple example from voting theory. Anthony Downs (1957) claimed that a rational voter (supposedly someone who values her time and resources) would weigh the cost of voting against the gains received from voting. These rewards are asserted to be the value from a preferred candidate winning the election times the probability that her vote will make an actual difference in the election. It is common to "measure" the difference between cost and reward as the **utility** that the person receives from the act. "Utility" is a word borrowed from economists that simply specifies an underlying preference scale that we usually cannot directly see. This is generally not complicated: I will get greater utility from winning the state lottery than I will from winning the office football pool, or I will get greater utility from spending time with my friends than I will from mowing the lawn.

Now we should make this idea more "mathematical" by specifying a relevant relationship. Riker and Ordeshook (1968) codified the Downsian model into mathematical symbology by articulating the following variables for an individual voter given a choice between two candidates:

R	=	the utility satisfaction of voting
P	=	the actual probability that the voter will affect the outcome with her particular vote
B	=	the perceived difference in benefits between the two candidates measured in utiles (units of utility): $B1 - B2$
C	=	the actual cost of voting in utiles (i.e., time, effort, money).

Thus the Downsian model is thus represented as

$$R = PB - C.$$

This is an unbelievably simple yet powerful model of political participation. In fact, we can use this statement to make claims that ***would not be as clear or as precise if described in descriptive language alone***. For instance, consider these statements:

- The voter will abstain if $R < 0$.
- The voter may still not vote even if $R > 0$ if there exist other competing activities that produce a higher R.
- If P is very small (i.e., it is a large election with many voters), then it is unlikely that this individual will vote.

The last statement leads to what is called ***the paradox of participation***: If nobody's vote is decisive, then why would anyone vote? Yet we can see that many people actually do show up and vote in large general elections. This paradox demonstrates that there is more going on than our simple model above.

The key point from the example above is that the formalism such mathematical representation provides gives us a way to say more exact things about social phenomena. Thus the motivation for introducing mathematics into the study of the social and behavioral sciences is to aid our understanding and improve the way we communicate substantive ideas.

1.4 Basic Terms

Some terms are used ubiquitously in social science work. A **variable** is just a ***symbol*** that represents a single number or group of numbers. Often variables are used as a substitution for numbers that we do not know or numbers that we will soon observe from some political or social phenomenon. Most frequently these are quantities like X, Y, a, b, and so on. Oddly enough, the modern notion of a variable was not codified until the early nineteenth century by the German mathematician Lejeune Dirichlet. We also routinely talk about **data**: collections of observed phenomenon. Note that ***data*** is plural; a single point is called a ***datum*** or a ***data point***.

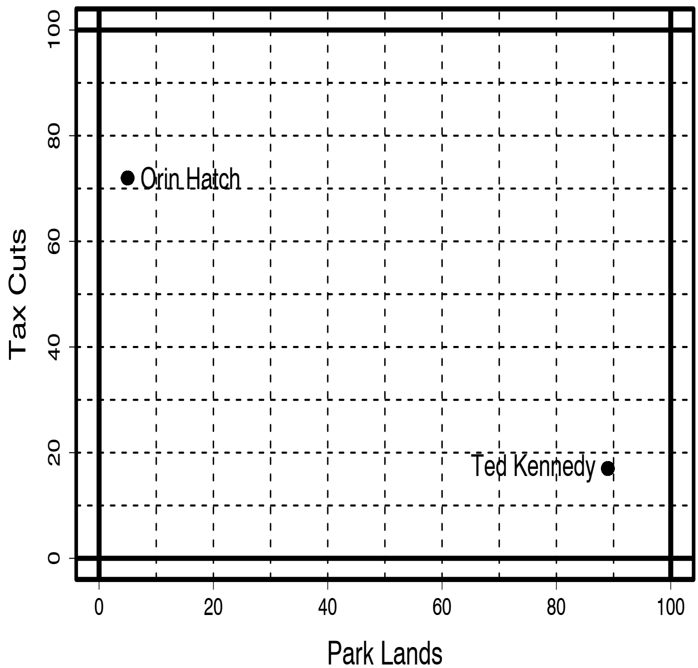
There are some other conventions from mathematics and statistics (as well as some other fields) that are commonly used in social science research as well. Some of these are quite basic, and social scientists speak this technical language fluently. Unless otherwise stated, variables are assumed to be defined on the **Cartesian coordinate system**.† If we are working with two variables x and y, then there is an assumed perpendicular set of axes where the x-axis (always given horizontally) is crossed with the y-axis (usually given vertically), such that the number pair (x, y) defines a **point** on the two-dimensional graph. There is actually no restriction to just two dimensions; for instance a point in 3-space is typically notated (x, y, z).

★ **Example 1.2: Graphing Ideal Points in the Senate**. One very active area of empirical research in political science is the estimation and subsequent use of ***legislative ideal points*** [see Jackman (2001), Londregan (2000), Poole and Rosenthal (1985, 1997)]. The objective is to analyze a member's voting record with the idea that this member's ideal policy position in policy-space can be estimated. This gets really interesting when the entire chamber (House, Senate, Parliament) is estimated accordingly, and various voting outcomes are analyzed or predicted.

Figure 1.1 shows approximate ideal points for Ted Kennedy and Oren Hatch on two proposed projects (it is common to propose Hatch as the foil for Kennedy). Senator Hatch is assumed to have an ideal point in this two-dimensional space at $x = 5, y = 72$, and Ted Kennedy is assumed to have an ideal point at $x = 89, y = 17$. These values are obtained from interest group rankings provided by the League of Conservation voters (2003) and the National Taxpayers Union (2003). We can also estimate the ideal points of other Senators in this way: One would guess that Trent Lott would be closer to Hatch than Kennedy, for instance.

† Alternatives exist such as "spherical space," where lines are defined on a generalization of circular space so they cannot be parallel to each other and must return to their point of origin, as well as Lobachevskian geometry and Kleinian geometry. These and other related systems are not generally useful in the social sciences and will therefore not be considered here with the exception of general polar coordinates in Chapter 2.

Fig. 1.1. Two Ideal Points in the Senate



Now consider a hypothetical trade-off between two bills competing for limited federal resources. These are appropriations (funding) for new national park lands, and a tax cut (i.e., national resources protection and development versus reducing taxes and thus taking in less revenue for the federal government). If there is a greater range of possible compromises, then other in-between points are possible. The best way to describe the possible space of solutions here is on a two-dimensional Cartesian coordinate system. Each Senator is assumed to have an ***ideal*** spending level for the two projects that trades off spending in one dimension against another: the level he or she would pick if they controlled the Senate completely. By convention we bound this in the two dimensions from 0 to 100.

The point of Figure 1.1 is to show how useful the Cartesian coordinate system is at describing positions along political and social variables. It might be more crowded, but it would not be more complicated to map the entire Senate along these two dimensions. In cases where more dimensions are considered, the graphical challenges become greater. There are two choices: show a subset on a two- or three-dimensional plot, or draw combinations of dimensions in a two-dimensional format by pairing two at a time.

Actually, in this Senate example, the use of the Cartesian coordinate system has been made quite restrictive for ease of analysis in this case. In the more typical, and more general, setting both the x-axis and the y-axis span negative infinity to positive infinity (although we obviously cannot ***draw*** them that way), and the space is labeled $\Re^2$ to denote the crossing of two ***real lines***. The **real line** is the line from minus infinity to positive infinity that contains the real numbers: numbers that are expressible in fractional form ($2/5, 1/3$, etc.) as well as those that are not because they have nonrepeating and infinitely continuing decimal values. There are therefore an infinite quantity of real numbers for any interval on the real line because numbers like $\sqrt{2}$ exist without "finishing" or repeating patterns in their list of values to the right of the decimal point ($\sqrt{2} = 1.4142135623730950488016887242096980785696718753769480731 7\ldots$).

It is also common to define various sets along the real line. These sets can be convex or nonconvex. A **convex set** has the property that for any two members of the set (numbers) x_1 and x_2, the number $x_3 = \delta x_1 + (1-\delta)x_2$ (for $0 \leq \delta \leq 1$) is also in the set. For example, if $\delta = \frac{1}{2}$, then x_3 is the average (the mean, see below) of x_1 and x_2.

In the example above we would say that Senators are constrained to express their preferences in the interval $[0\!:\!100]$, which is commonly used as a measure of ideology or policy preference by interest groups that rate elected officials [such as the ***Americans for Democratic Action*** (ADA), and the ***American Conservative Union*** (ACU)]. **Interval notation** is used frequently in mathematical notation, and there is only one important distinction: Interval ends can be "open" or "closed." An open interval excludes the end-point denoted with parenthetical forms "(" and ")" whereas the closed interval denoted with bracket forms "[" and "]" includes it (the curved forms "{" and "}" are usually reserved for set notation). So, in altering our Senate example, we have the following one-dimensional options for x (also for y):

open on both ends:	$(0{:}100)$,	$0 < x < 100$
closed on both ends:	$[0{:}100]$,	$0 \le x \le 100$
closed left, open right	$[0{:}100)$,	$0 \le x < 100$
open left, closed right	$(0{:}100]$,	$0 < x \le 100$

Thus the restrictions on δ above are that it must lie in $[0{:}1]$. These intervals can also be expressed in ***comma notation*** instead of ***colon notation***: $[0, 100]$.

1.4.1 Indexing and Referencing

Another common notation is the technique of indexing observations on some variable by the use of subscripts. If we are going to list some value like years served in the House of Representatives (as of 2004), we would not want to use some cumbersome notation like

$$\begin{array}{rcc}
\text{Abercrombie} = 14 & \qquad & \vdots \\
\text{Acevedo-Vila} = 14 & & \text{Wu} = 6 \\
\text{Ackerman} = 21 & & \text{Wynn} = 12 \\
\text{Aderholt} = 8 & & \text{Young} = 34 \\
\vdots & & \text{Young} = 32
\end{array}$$

which would lead to awkward statements like "Abercrombie's years in office" + "Acevedo-Vila's years in office"... + "Young's years in office" to express mathematical manipulation (note also the obvious naming problem here as well, i.e., delineating between Representative Young of Florida and Representative Young of Alaska). Instead we could just assign each member ordered alphabetically to an integer 1 through 435 (the number of U.S. House members) and then index them by subscript: $\mathbf{X} = \{X_1, X_2, X_3, \ldots, X_{433}, X_{434}, X_{435}\}$. This is a lot cleaner and more mathematically useful. For instance, if we wanted to calculate the mean (average) time served, we could simply perform:

$$\overline{\mathbf{X}} = \frac{1}{435}\left(X_1 + X_2 + X_3 + \cdots + X_{433} + X_{434} + X_{435}\right)$$

(the bar over $\mathbf{X}$ denotes that this average is a ***mean***, something we will see frequently). Although this is cleaner and easier than spelling names or something like that, there is an even nicer way of indicating a mean calculation that uses the **summation operator**. This is a large version of the Greek letter sigma where the starting and stopping points of the addition process are spelled out over and under the symbol. So the mean House seniority calculation could be specified simply by

$$\bar{\mathbf{X}} = \frac{1}{435}\sum_{i=1}^{435} X_i,$$

where we say that i ***indexes*** X in the summation. One way to think of this notation is that $\sum$ is just an adding "machine" that instructs us which X to start with and which one to stop with. In fact, if we set $n = 435$, then this becomes the simple (and common) form

$$\bar{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

More formally,

The Summation Operator

- $\rightarrowtail$ If $X_1, X_2, \ldots, X_n$ are n numerical values,
- $\rightarrowtail$ then their sum can be represented by $\sum_{i=1}^{n} X_i$,
- $\rightarrowtail$ where i is an indexing variable to indicate the starting and stopping points in the series $X_1, X_2, \ldots, X_n$.

A related notation is the **product operator**. This is a slightly different "machine" denoted by an uppercase Greek pi that tells us to multiply instead of add as we did above:

$$\prod_{i=1}^{n} X_i$$

(i.e., it multiplies the n values together). Here we also use i again as the index, but it is important to note that there is nothing special about the use of i; it is just a very common choice. Frequent index alternatives include j, k, l, and m. As a simple illustration, suppose $p_1 = 0.2$, $p_2 = 0.7$, $p_3 = 0.99$, $p_4 = 0.99$, $p_5 = 0.99$. Then

$$\prod_{j=1}^{5} p_j = p_1 \cdot p_2 \cdot p_3 \cdot p_4 \cdot p_5$$
$$= (0.2)(0.7)(0.99)(0.99)(0.99)$$
$$= 0.1358419.$$

Similarly, the formal definition for this operator is given by

The Product Operator

- ↣ If $X_1, X_2, \ldots, X_n$ are n numerical values,
- ↣ then their product can be represented by $\prod_{i=1}^{n} X_i$,
- ↣ where i is an indexing variable to indicate the starting and stopping points in the series $X_1, X_2, \ldots, X_n$.

Subscripts are used because we can immediately see that they are not a mathematical operation on the symbol being modified. Sometimes it is also convenient to index using a superscript. To distinguish between a superscript as an index and an exponent operation, brackets or parentheses are often used. So X^2 is the square of X, but $X^{[2]}$ and $X^{(2)}$ are indexed values.

There is another, sometimes confusing, convention that comes from six decades of computer notation in the social sciences and other fields. Some authors will index values without the subscript, as in $X1, X2, \ldots$, or differing functions (see Section 1.5 for the definition of a function) without subscripting according to $f1, f2, \ldots$. Usually it is clear what is meant, however.

1.4.2 Specific Mathematical Use of Terms

The use of mathematical terms can intimidate readers even when the author does not mean to do so. This is because many of them are based on the Greek alphabet or strange versions of familiar symbols (e.g., $\forall$ versus A). This does not mean that the use of these symbols should be avoided for readability. Quite the opposite; for those familiar with the basic vocabulary of mathematics such symbols provide a more concise and readable story if they can clearly summarize ideas that would be more elaborate in narrative. We will save the complete list of Greek idioms to the appendix and give others here, some of which are critical in forthcoming chapters and some of which are given for completeness.

Some terms are almost universal in their usage and thus are important to recall without hesitation. Certain probability and statistical terms will be given as needed in later chapters. An important group of standard symbols are those that define the set of numbers in use. These are

Symbol	Explanation
$\mathfrak{R}$	the set of real numbers
$\mathfrak{R}^+$	the set of positive real numbers
$\mathfrak{R}^-$	the set of negative real numbers
$\mathcal{I}$	the set of integers
$\mathcal{I}^+$ or $\mathbb{Z}^+$	the set of positive integers
$\mathcal{I}^-$ or $\mathbb{Z}^+$	the set of negative integers
$\mathbf{Q}$	the set of rational numbers
$\mathbf{Q}^+$	the set of positive rational numbers
$\mathbf{Q}^-$	the set of negative rational numbers
$\mathcal{C}$	the set of complex numbers (those based on $\sqrt{-1}$).

Recall that the real numbers take on an infinite number of values: rational (expressible in fraction form) and irrational (not expressible in fraction form with values to the right of the decimal point, nonrepeating, like pi). It is interesting to note that there are an infinite number of irrationals and every irrational falls between two rational numbers. For example, $\sqrt{2}$ is in between $7/5$ and $3/2$. Integers are positive and negative (rational) numbers with no decimal component and sometimes called the "counting numbers." Whole numbers are positive integers along with zero, and natural numbers are positive integers without zero. We will not generally consider here the set of **complex numbers**, but they are those that include the imaginary number: $i = \sqrt{-1}$, as in $\sqrt{-4} = 2\sqrt{-1} = 2i$. In mathematical and statistical modeling it is often important to remember which of these number types above is being considered.

Some terms are general enough that they are frequently used with statements about sets or with standard numerical declarations. Other forms are more obscure but do appear in certain social science literatures. Some reasonably common examples are listed in the next table. Note that all of these are ***contextual***, that is, they lack any meaning outside of sentence-like statements with other symbols.

Symbol	Explanation
$\neg$	logical negation statement
$\in$	is an element of, as in $3 \in \mathcal{I}^+$
$\ni$	such that
$\therefore$	therefore
$\because$	because
$\Longrightarrow$	logical "then" statement
$\Longleftrightarrow$	if and only if, also abbreviated "iff"
$\exists$	there exists
$\forall$	for all
$\between$	between
$\parallel$	parallel
$\angle$	angle

Also, many of these symbols can be negated, and negation is expressed in one of two ways. For instance, $\in$ means "is an element of," but both $\notin$ and $\neg \in$ mean "is ***not*** an element of." Similarly, $\subset$ means "is a subset of," but $\not\subset$ means "is ***not*** a subset of."

Some of these terms are used in a very linguistic fashion: $3-4 \in \Re^- \because 3 < 4$. The "therefore" statement is usually at the end of some logic: $2 \in \mathcal{I}^+ \therefore 2 \in \Re^+$. The last three in this list are most useful in geometric expressions and indicate spatial characteristics. Here is a lengthy mathematical statement using most of these symbols: $\forall x \in \mathcal{I}^+$ and $x\neg$prime, $\exists y \in \mathcal{I}^+ \ni x/y \in \mathcal{I}^+$. So what does this mean? Let's parse it: "For all numbers x such that x is a positive integer and not a prime number, there exists a y that is a positive integer such that x divided by y is also a positive integer." Easy, right? (Yeah, sure.) Can you construct one yourself?

Another "fun" example is $x \in \mathcal{I}$ and $x \neq 0 \implies x \in \mathcal{I}^-$ or $\mathcal{I}^+$. This says that if x is a nonzero integer, it is either a positive integer or a negative integer. Consider this in pieces. The first part, $x \in \mathcal{I}$, stipulates that x is "in" the group of integers and cannot be equal to zero. The right arrow, $\implies$, is a logical consequence statement equivalent to saying "then." The last part gives the result, either x is a negative integer or a positive integer (and nothing else since no alternatives are given).

Another important group of terms are related to the manipulation of ***sets*** of objects, which is an important use of mathematics in social science work (sets are simply defined groupings of individual objects; see Chapter 7, where sets and operations on sets are defined in detail). The most common are

Symbol	Explanation
$\emptyset$	the empty set (sometimes used with the Greek phi: ϕ)
$\cup$	union of sets
$\cap$	intersection of sets
$\setminus$	subtract from set
$\subset$	subset
$\complement$	complement

These allow us to make statements about groups of objects such as $A \subset B$ for $A = \{2, 4\}$, $B = \{2, 4, 7\}$, meaning that the set A is a smaller grouping of the larger set B. We could also observe that the A results from removing seven from B.

Some symbols, however, are "restricted" to comparing or operating on strictly numerical values and are not therefore applied directly to sets or logic expressions. We have already seen the sum and product operators given by the symbols $\sum$ and $\prod$ accordingly. The use of ∞ for infinity is relatively common even outside of mathematics, but the next list also gives two distinct "flavors" of infinity. Some of the contexts of these symbols we will leave to remaining chapters as they deal with notions like limits and vector quantities.

Symbol	Explanation
$\propto$	is proportional to
$\doteq$	equal to in the limit (approaches)
$\perp$	perpendicular
∞	infinity
∞^+, $+\infty$	positive infinity
∞^-, $-\infty$	negative infinity
$\sum$	summation
$\prod$	product
$\lfloor\ \rfloor$	floor: round down to nearest integer
$\lceil\ \rceil$	ceiling: round up to nearest integer
$\vert$	given that: $X\vert Y = 3$

Related to these is a set of functions relating maximum and minimum values. Note the directions of $\vee$ and $\wedge$ in the following table.

Symbol	Explanation
$\vee$	maximum of two values
$\max()$	maximum value from list
$\wedge$	minimum of two values
$\min()$	minimum value from list
$\underset{x}{\text{argmax}} f(x)$	the value of x that maximizes the function $f(x)$
$\underset{x}{\text{argmin}} f(x)$	the value of x that minimizes the function $f(x)$

The latter two are important but less common functions. Functions are formally defined in the next section, but we can just think of them for now as sets of instructions for modifying input values (x^2 is an example function that squares its input). As a simple example of the argmax function, consider

$$\underset{x \in \Re}{\text{argmax}}\ x(1-x),$$

which asks which value on the real number line maximizes $x(1-x)$. The answer is 0.5 which provides the best trade-off between the two parts of the function. The argmin function works accordingly but (obviously) operates on the function minimum instead of the function maximum.

These are not exhaustive lists of symbols, but they are the most fundamental (many of them are used in subsequent chapters). Some literatures develop their own conventions about symbols and their very own symbols, such as $\Game$ to denote a mathematical representation of a game and $\Bumpeq$ to indicate ***geometric*** equivalence between two objects, but such extensions are rare in the social sciences.

1.5 Functions and Equations

A mathematical equation is a very general idea. Fundamentally, an **equation** "equates" two quantities: They are arithmetically identical. So the expression $R = PB - C$ is an equation because it establishes that R and $PB - C$ are exactly equal to each other. But the idea of a mathematical sentence is more general (less restrictive) than this because we can substitute other relations for equality, such as

Symbol	Meaning
$<$	less than
$\leq$	less than or equal to
$\ll$	much less than
$>$	greater than
$\geq$	greater than or equal to
$\gg$	much greater than
$\approx$	approximately the same
$\cong$	approximately equal to
$\lessapprox$	approximately less than (also $\lesssim$)
$\gtrapprox$	approximately greater than (also $\gtrsim$)
$\equiv$	equivalent by assumption

So, for example, if we say that $X = 1$, $Y = 1.001$ and $Z = 0.002$, then the following statements are true:

$$X \leq 1 \qquad X \geq 1 \qquad X \ll 1000 \qquad X \gg -1000$$

$$X < 2 \qquad X > 0 \qquad X \cong 0.99. \qquad X \approx 1.0001X$$

$$X \approx Y \qquad Y \lessapprox X + Z \qquad X + Z \gtrapprox Y$$

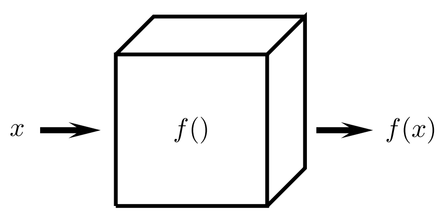
$$X + 0.001 \equiv Y \qquad X > Y - Z \qquad X \propto 2Y.$$

The purpose of the equation form is generally to express more than one set of relations. Most of us remember the task of solving "two equations for two unknowns." Such forms enable us to describe how (possibly many) variables are associated to each other and various constants. The formal language of mathematics relies heavily on the idea that equations are the atomic units of relations.

What is a function? A **mathematical function** is a "mapping" (i.e., specific directions), which gives a correspondence from one measure onto exactly one other for that value. That is, in our context it defines a relationship between one variable on the x-axis of a Cartesian coordinate system and an operation on that variable that can produce only one value on the y-axis. So a function is a ***mapping*** from one defined space to another, such as $f : \mathfrak{R} \rightarrow \mathfrak{R}$, in which f maps the real numbers to the real numbers (i.e., $f(x) = 2x$), or $f : \mathfrak{R} \rightarrow \mathcal{I}$, in which f maps the real numbers to the integers (i.e., $f(x) = \text{round}(x)$).

This all sounds very technical, but it is not. One way of thinking about functions is that they are a "machine" for transforming values, sort of a box as in the figure to the right.

A Function Represented



To visualize this we can think about values, x, going in and some modification of these values, $f(x)$, coming out where the instructions for this process are

Table 1.1. Tabularizing $f(x) = x^2 - 1$

x	$f(x) = x^2 - 1$
1	0
3	8
-1	0
10	99
4	15
$\sqrt{3}$	2

contained in the "recipe" given by $f()$.

Consider the following function operating on the variable x:

$$f(x) = x^2 - 1.$$

This simply means that the mapping from x to $f(x)$ is the process that squares x and subtracts 1. If we list a set of inputs, we can define the corresponding set of outputs, for example, the paired values listed in Table 1.1.

Here we used the $f()$ notation for a function (first codified by Euler in the eighteenth century and still the most common form used today), but other forms are only slightly less common, such as: $g()$, $h()$, $p()$, and $u()$. So we could have just as readily said:

$$g(x) = x^2 - 1.$$

Sometimes the additional notation for a function is essential, such as when more than one function is used in the same expression. For instance, functions can be "nested" with respect to each other (called a composition):

$$f \circ g = f(g(x)),$$

as in $g(x) = 10x$ and $f(x) = x^2$, so $f \circ g = (10x)^2$ (note that this is different than $g \circ f$, which would be $10(x^2)$). Function definitions can also contain wording instead of purely mathematical expressions and may have conditional aspects. Some examples are

$$f(y) = \begin{cases} \frac{1}{y} \text{ if } y \neq 0 \text{ and } y \text{ is rational} \\ 0 \text{ otherwise} \end{cases}$$

$$p(x) = \begin{cases} (6-x)^{-\frac{5}{3}}/200 + 0.1591549 & \text{for } \theta \in [0{:}6) \\ \frac{1}{2\pi}\frac{1}{\left(1+\left(\frac{x-6}{2}\right)^2\right)} & \text{for } \theta \in [6{:}12]. \end{cases}$$

Note that the first example is necessarily a noncontinuous function whereas the second example is a continuous function (but perhaps not obviously so). Recall that π is notation for $3.1415926535\ldots$, which is often given inaccurately as just 3.14 or even $22/7$. To be more specific about such function characteristics, we now give two important properties of a function.

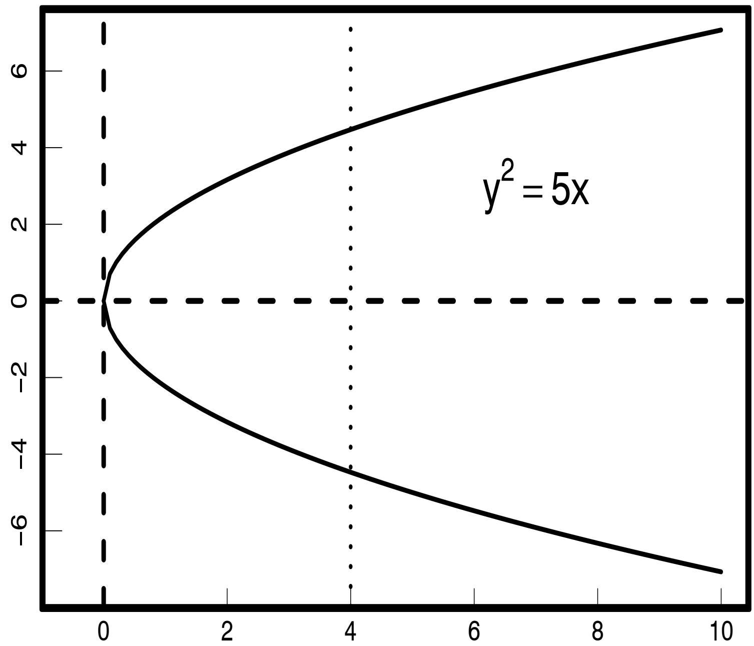
Properties of Functions, Given for $g(x) = y$

$\rightarrowtail$ A function is **continuous** if it has no “gaps” in its mapping from x to y.

$\rightarrowtail$ A function is **invertible** if its reverse operation exists: $g^{-1}(y) = x$, where $g^{-1}(g(x)) = x$.

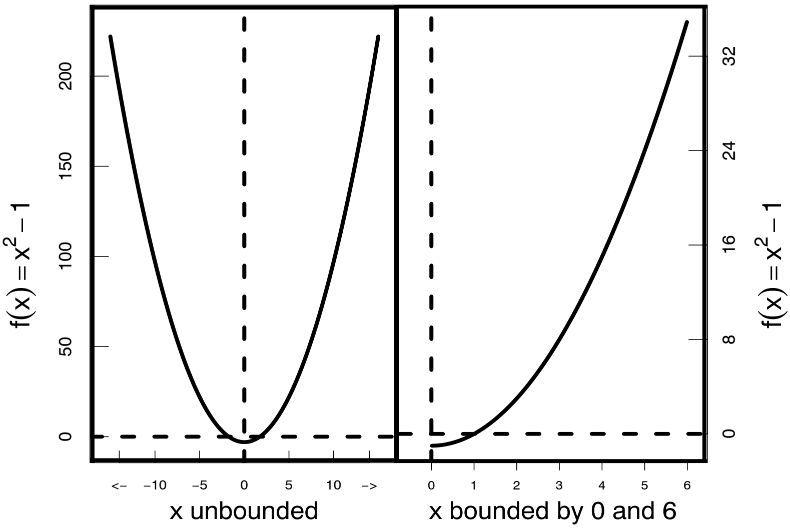
It is important to distinguish between a function and a **relation**. A function must have ***exactly one value returned by*** $f(x)$ ***for each value of*** x, whereas a relation does not have this restriction. One way to test whether $f(x)$ is a function or, more generally, a relation is to graph it in the Cartesian coordinate system (x versus y in orthogonal representation) and see if there is a vertical line that can be drawn such that it intersects the function at two values (or more) of y for a single value of x. If this occurs, then it is not a function. There is an important distinction to be made here. The ***solution*** to a function can possibly have more than one corresponding value of x, but a function cannot have alternate values of y for a ***given*** x. For example, consider the relation $y^2 = 5x$, which is not a function based on this criteria. We can see this algebraically by taking the square root of both sides, $\pm y = \sqrt{5x}$, which shows the non-uniqueness of the y values (as well as the restriction to positive values of x). We can also see this graphically in Figure 1.2, where x values from 0 to 10 each give two y values (a dotted line is given at $(x = 4, y = \pm\sqrt{20})$ as an example).

Fig. 1.2. A Relation That Is Not a Function



The modern definition of a function is also attributable to Dirichlet: If variables x and y are related such that every acceptable value of x has a corresponding value of y defined by a ***rule***, then y is a function of x. Earlier European period notions of a function (i.e., by Leibniz, Bernoulli , and Euler) were more vague and sometimes tailored only to specific settings.

Fig. 1.3. RELATING x AND $f(x)$



Often a function is explicitly defined as a mapping between elements of an **ordered pair**: (x, y), also called a relation. So we say that the function $f(x) = y$ maps the ordered pair x, y such that for each value of x there is exactly one y (the order of x before y matters). This was exactly what we saw in Table 1.1, except that we did not label the rows as ordered pairs. As a more concrete example, the following set of ordered pairs:

$$\{[1, -2],\ [3, 6],\ [7, 46]\}$$

can be mapped by the function: $f(x) = x^2 - 3$. If the set of x values is restricted to some specifically defined set, then obviously so is y. The set of x values is called the **domain** (or support) of the function and the associated set of y values is called the **range** of the function. Sometimes this is highly restrictive (such as to specific integers) and sometimes it is not. Two examples are given in Figure 1.3, which is drawn on the (now) familiar Cartesian coordinate system. Here we see that the range and domain of the function are unbounded in the first panel (although we clearly cannot ***draw*** it all the way until infinity in both directions), and the domain is bounded by 0 and 6 in the second panel.

A function can also be even or odd, defined by

$$\text{a function is "odd" if:} \quad f(-x) = -f(x)$$
$$\text{a function is "even" if:} \quad f(-x) = f(x).$$

So, for example, the squaring function $f(x) = x^2$ and the absolute value function $f(x) = |x|$ are even because both will always produce a positive answer. On the other hand, $f(x) = x^3$ is odd because the negative sign perseveres for a negative x. Regretfully, functions can also be neither even nor odd without domain restrictions.

One special function is important enough to mention directly here. A **linear function** is one that preserves the algebraic nature of the real numbers such that $f()$ is a linear function if:

$$f(x_1 + x_2) = f(x_1) + f(x_2) \qquad \text{and} \quad f(kx_1) = kf(x_1)$$

for two points, x_1 and x_2, in the domain of $f()$ and an arbitrary constant number k. This is often more general in practice with multiple functions and multiple constants, forms such as:

$$F(x_1, x_2, x_3) = kf(x_1) + \ell g(x_2) + mh(x_3)$$

for functions $f(), g(), h()$ and constants k, ℓ, m.

★ **Example 1.3: The "Cube Rule" in Votes to Seats**. A standard, though somewhat maligned, theory from the study of elections is due to Parker's (1909) empirical research in Britain, which was later popularized in that country by Kendall and Stuart (1950, 1952). He looked at systems with two major parties whereby the largest vote-getter in a district wins regardless of the size of the winning margin (the so-called ***first past the post*** system used by most English-speaking countries). Suppose that A denotes the proportion of votes for one party and B the proportion of votes for the other. Then, according to this rule, the ratio of seats in Parliament won is approximately the cube of the ratio of votes: A/B in votes implies A^3/B^3 in seats (sometimes ratios are given in the notation $A:B$). The political principle from this theory is that small differences in the vote ratio yield large differences in the seats ratio and thus provide stable parliamentary government.

So how can we express this theory in standard mathematical function notation. Define x as the ratio of votes for the party with proportion A over the party with proportion B. Then expressing the cube law in this notation yields

$$f(x) = x^3$$

for the function determining seats, which of course is very simple. Tufte (1973) reformulated this slightly by noting that in a two-party contest the proportion of votes for the second party can be rewritten as $B = 1 - A$. Furthermore, if we define the proportion of ***seats*** for the first party as S_A, then similarly the proportion of seats for the second party is $1 - S_A$, and we can reexpress the cube rule in this notation as

$$\frac{S_A}{1 - S_A} = \left[\frac{A}{1 - A}\right]^3.$$

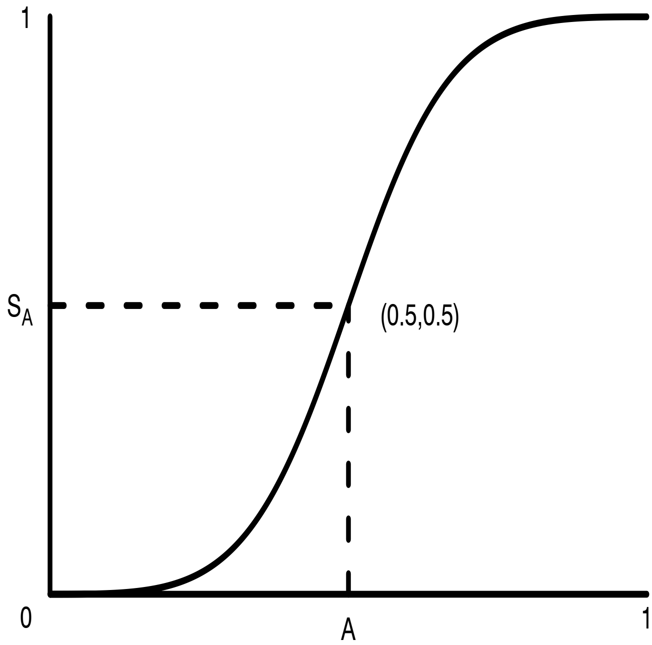
Using this notation we can solve for S_A (see Exercise 1.8), which produces

$$S_A = \frac{A^3}{1 - 3A + 3A^2}.$$

This equation has an interesting shape with a rapid change in the middle of the range of A, clearly showing the nonlinearity in the relationship implied by the cube function. This shape means that the winning party's gains are more pronounced in this area and less dramatic toward the tails. This is shown in Figure 1.4.

Taagepera (1986) looked at this for a number of elections around the world and found some evidence that the rule fits. For instance, U.S. House races for the period 1950 to 1970 with Democrats over Republicans give a value of exactly 2.93, which is not too far off the theoretical value of 3 supplied by Parker.

Fig. 1.4. The Cube Law



1.5.1 Applying Functions: The Equation of a Line

Recall the familiar expression of a line in Cartesian coordinates usually given as $y = mx + b$, where m is the slope of the line (the change in y for a one-unit change in x) and b is the point where the line intercepts the y-axis. Clearly this is a (linear) function in the sense described above and also clearly we can determine any single value of y for a given value of x, thus producing a matched pair.

A classic problem is to find the slope and equation of a line determined by two points. This is always unique because any two points in a Cartesian coordinate system can be connected by one and only one line. Actually we can generalize this in a three-dimensional system, where three points determine a unique plane, and so on. This is why a three-legged stool never wobbles and a four-legged chair sometimes does (think about it!). Back to our problem…suppose that we want to find the equation of the line that goes through the two points

$\{[2,1],[3,5]\}$. What do we know from this information? We know that for one unit of increasing x we get four units of increasing y. Since slope is "rise over run," then:

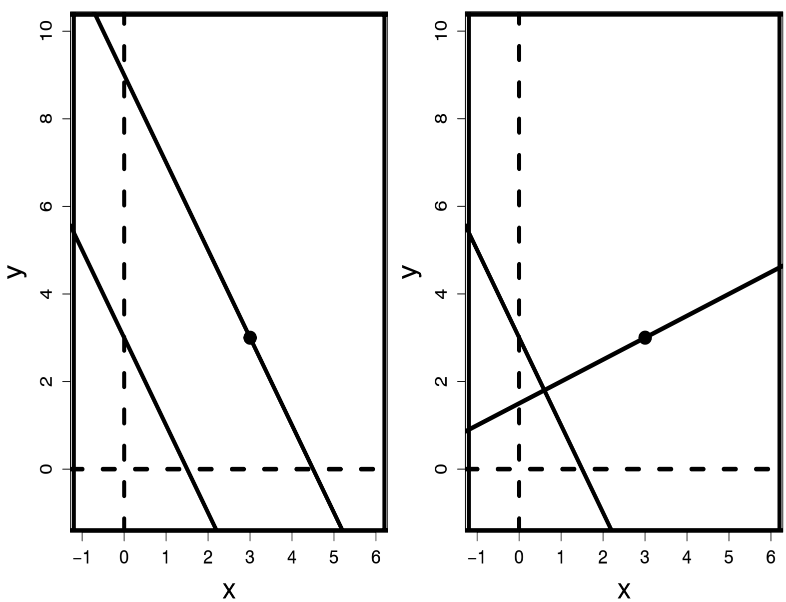
$$m = \frac{5-1}{3-2} = 4.$$

Great, now we need to get the intercept. To do this we need only to plug m into the standard line equation, set x and y to one of the known points on the line, and solve (we should pick the easier point to work with, by the way):

$$y = mx + b$$
$$1 = 4(2) + b$$
$$b = 1 - 8 = -7.$$

This is equivalent to starting at some selected point on the line and "walking down" until the point where x is equal to zero.

Fig. 1.5. Parallel and Perpendicular Lines



The Greeks and other ancients were fascinated by linear forms, and lines are an interesting mathematical subject unto themselves. For instance, two lines

$$y = m_1 x + b_1$$

$$y = m_2 x + b_2,$$

are parallel ***if and only if*** (often abbreviated as "iff") $m_1 = m_2$ and perpendicular (also called orthogonal) iff $m_1 = -1/m_2$. For example, suppose we have the line $L_1 : y = -2x + 3$ and are interested in finding the line parallel to L_1 that goes through the point $[3, 3]$. We know that the slope of this new line must be -2, so we now plug this value in along with the only values of x and y that we know are on the line. This allows us to solve for b and plot the parallel line in left panel of Figure 1.5:

$$(3) = -2(3) + b_2, \qquad \text{so} \qquad b_2 = 9.$$

This means that the parallel line is given by $L_2 : y = -2x + 9$. It is not much more difficult to get the equation of the perpendicular line. We can do the same trick but instead plug in the negative inverse of the slope from L_1:

$$(3) = \frac{1}{2}(3) + b_3, \qquad \text{so} \qquad b_3 = \frac{3}{2}.$$

This gives us $L_2 \perp L_1$, where $L_2 : y = \frac{1}{2}x + \frac{3}{2}$.

★ **Example 1.4: Child Poverty and Reading Scores**. Despite overall national wealth, a surprising number of U.S. school children live in poverty. A continuing concern is the effect that this has on educational development and attainment. This is important for normative as well as societal reasons. Consider the following data collected in 1998 by the California Department of Education (CDE) by testing all 2nd–11th grade students in various subjects (the Stanford 9 test). These data are aggregated to the school district level here for two variables: the percentage of students who qualify for reduced or free lunch plans (a common measure of poverty in educational policy studies) and the percent of students scoring over the national median for reading at the 9th grade. The median (average) is the point where one-half of the points are greater and one-half of the points are less.

Because of the effect of limited English proficiency students on district performance, this test was by far the most controversial in California amongst the exam topics. In addition, administrators are sensitive to the aggregated results of reading scores because it is a subject that is at the core of what many consider to be "traditional" children's education.

The relationship is graphed in Figure 1.6 along with a linear trend with a slope of $m = -0.75$ and an intercept at $b = 81$. A very common tool of social scientists is the so-called **linear regression model**. Essentially this is a method of looking at data and figuring out an underlying trend in the form of a straight line. We will not worry about any of the calculation details here, but we can think about the implications. What does this particular line mean? It means that for a 1% positive change (say from 50 to 51) in a district's poverty, they will have an ***expected*** reduction in the pass rate of three-quarters of a percent. Since this line purports to find the underlying trend across these 303 districts, no district will ***exactly*** see these results, but we are still claiming that this captures some common underlying socioeconomic phenomena.

1.5.2 The Factorial Function

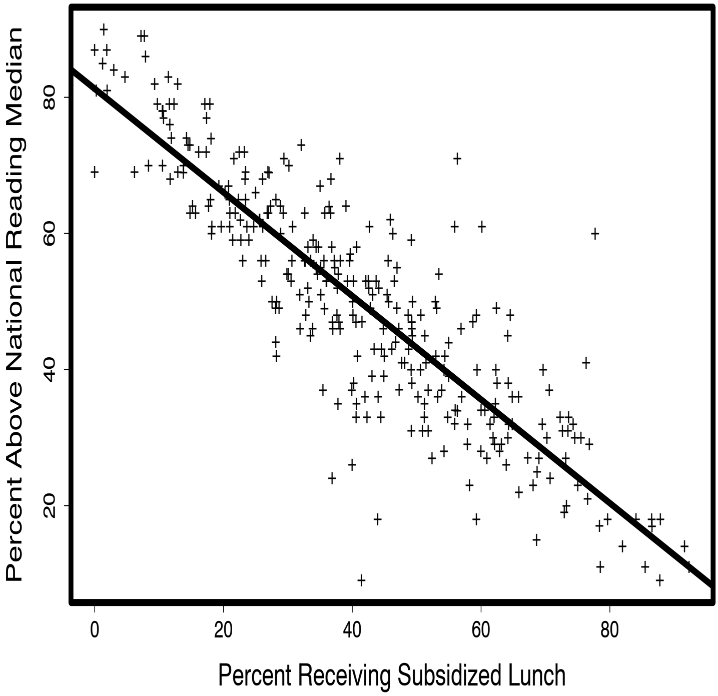
One function that has special notation is the factorial function. The factorial of x is denoted $x!$ and is defined for positive integers x only:

$$x! = x \times (x-1) \times (x-2) \times \ldots 2 \times 1,$$

where the 1 at the end is superfluous. Obviously $1! = 1$, and by convention we assume that $0! = 1$. For example,

$$4! = 4 \times 3 \times 2 \times 1 = 24.$$

Fig. 1.6. Poverty and Reading Test Scores



It should be clear that this function grows rapidly for increasing values of x, and sometimes the result overwhelms commonly used hand calculators. Try, for instance, to calculate $100!$ with yours. In some common applications large factorials are given in the context of ratios and a handy cancellation can be used to make the calculation easier. It would be difficult or annoying to calculate $190!/185!$ by first obtaining the two factorials and then dividing. Fortunately we can use

$$\begin{aligned}\frac{190!}{185!} &= \frac{190 \cdot 189 \cdot 188 \cdot 187 \cdot 186 \cdot 185 \cdot 184 \cdot 183 \cdot \ldots}{185 \cdot 184 \cdot 183 \cdot \ldots} \\ &= 190 \cdot 189 \cdot 188 \cdot 187 \cdot 186 \\ &= 234,816,064,560\end{aligned}$$

(recall that "$\cdot$" and "$\times$" are equivalent notations for multiplication). It would not initially seem like this calculation produces a value of almost 250 billion, but it does! Because factorials increase so quickly in magnitude, they can sometimes be difficult to calculate directly. Fortunately there is a handy way to get around this problem called **Stirling's Approximation** (curiously named since it is credited to De Moivre's 1720 work on probability):

$$n! \approx (2\pi n)^{\frac{1}{2}} e^{-n} n^n.$$

Here $e \approx 2.71$, which is an important constant defined on page 36. Notice that, as its name implies, this is an approximation. We will return to factorials in Chapter 7 when we analyze various counting rules.

★ **Example 1.5: Coalition Cabinet Formation**. Suppose we are trying to form a coalition cabinet with three parties. There are six senior members of the Liberal Party, five senior members of the Christian Democratic Party, and four senior members of the Green Party vying for positions in the cabinet. How many ways could you choose a cabinet composed of three Liberals, two Christian Democrats, and three Greens?

It turns out that the number of possible subsets of y items from a set of n items is given by the "choose notation" formula:

$$\binom{n}{y} = \frac{n!}{y!(n-y)!},$$

which can be thought of as the permutations of n divided by the permutations of y times the permutations of "not y." This is called ***unordered without replacement*** because it does not matter what order the members are drawn in, and once drawn they are not thrown back into the pool for possible re-selection. There are actually other ways to select samples from populations, and these are given in detail in Chapter 7 (see, for instance, the discussion in Section 7.2).

So now we have to multiply the number of ways to select three Liberals, the two CDPs, and the three Greens to get the ***total*** number of possible cabinets (we multiply because we want the full number of combinatoric possibilities across the three parties):

$$\binom{6}{3}\binom{5}{2}\binom{4}{3} = \frac{6!}{3!(6-3)!}\frac{5!}{2!(5-2)!}\frac{4!}{3!(4-3)!}$$
$$= \frac{720}{6(6)}\frac{120}{2(6)}\frac{24}{6(1)}$$
$$= 20 \times 10 \times 4$$
$$= 800.$$

This number is relatively large because of the multiplication: For each single choice of members from one party we have to consider ***every*** possible choice from the others. In a practical scenario we might have many fewer ***politically viable*** combinations due to overlapping expertise, jealousies, rivalries, and other interesting phenomena.

1.5.3 The Modulo Function

Another function that has special notation is the **modulo function**, which deals with the ***remainder*** from a division operation. First, let's define a **factor**: y is a factor of x if the result of x/y is an integer (i.e., a prime number has exactly two factors: itself and one). So if we divided x by y and y was ***not*** a factor of x, then there would necessarily be a noninteger remainder between zero and one. This remainder can be an inconvenience where it is perhaps discarded, or it can be considered important enough to keep as part of the result. Suppose instead that this was the only part of the result from division that we cared about. What symbology could we use to remove the integer component and only keep the remainder?

To divide x by y and keep only the remainder, we use the notation

$$x \pmod{y}.$$

Thus $5 \pmod{2} = 1$, $17 \pmod{5} = 2$, and $10,003 \pmod{4} = 3$, for example. The modulo function is also sometimes written as either

$$x \bmod y \qquad \text{or} \qquad x \mod y$$

(only the spacing differs).

1.6 Polynomial Functions

Polynomial functions of x are functions that have components that raise x to some power:

$$f(x) = x^2 + x + 1$$
$$g(x) = x^5 - 3^3 - x$$
$$h(x) = x^{100},$$

where these are polynomials in x of power 2, 5, and 100, respectively. We have already seen examples of polynomial functions in this chapter such as $f(x) = x^2$, $f(x) = x(1-x)$, and $f(x) = x^3$. The convention is that a polynomial degree (power) is designated by its largest exponent with regard to the variable. Thus the polynomials above are of degree 2,5, and 100, respectively.

Often we care about the **roots** of a polynomial function: where the curve of the function crosses the x-axis. This may occur at more than one place and may be difficult to find. Since $y = f(x)$ is zero at the x-axis, root finding means discovering where the right-hand side of the polynomial function equals zero. Consider the function $h(x) = x^{100}$ from above. We do not have to work too hard to find that the only root of this function is at the point $x = 0$.

In many scientific fields it is common to see **quadratic** polynomials, which are just polynomials of degree 2. Sometimes these polynomials have easy-to-determine integer roots (solutions), as in

$$x^2 - 1 = (x-1)(x+1) \Longrightarrow x = \pm 1,$$

and sometimes they do not, requiring the well-known quadratic equation

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

where a is the multiplier on the x^2 term, b is the multiplier on the x term, and c is the constant. For example, solving for roots in the equation

$$x^2 - 4x = 5$$

is accomplished by

$$x = \frac{-(-4) \pm \sqrt{(-4)^2 - 4(1)(-5)}}{2(1)} = -1 \text{ or } 5,$$

where $a = 1$, $b = -4$, and $c = -5$ from $f(x) = x^2 - 4x - 5 \equiv 0$.

1.7 Logarithms and Exponents

Exponents and logarithms ("logs" for short) confuse many people. However, they are such an important convenience that they have become critical to quantitative social science work. Furthermore, so many statistical tools use these "natural" expressions that understanding these forms is essential to some work. Basically exponents make convenient the idea of multiplying a number by itself (possibly) many times, and a logarithm is just the opposite operation. We already saw one use of exponents in the discussion of the cube rule relating votes to seats. In that example, we defined a function, $f(x) = x^3$, that used 3 as an exponent. This is only mildly more convenient than $f(x) = x \times x \times x$, but imagine if the exponent was quite large or if it was not a integer. Thus we need some core principles for handling more complex exponent forms.

First let's review the basic rules for exponents. The important ones are as follows.

Key Properties of Powers and Exponents

$\rightarrowtail$	Zero Property	$x^0 = 1$
$\rightarrowtail$	One Property	$x^1 = x$
$\rightarrowtail$	Power Notation	$\text{power}(x, a) = x^a$
$\rightarrowtail$	Fraction Property	$\left(\frac{x}{y}\right)^a = \left(\frac{x^a}{y^a}\right) = x^a y^{-a}$
$\rightarrowtail$	Nested Exponents	$(x^a)^b = x^{ab}$
$\rightarrowtail$	Distributive Property	$(xy)^a = x^a y^a$
$\rightarrowtail$	Product Property	$x^a \times x^b = x^{a+b}$
$\rightarrowtail$	Ratio Property	$x^{\frac{a}{b}} = (x^a)^{\frac{1}{b}} = \left(x^{\frac{1}{b}}\right)^a = \sqrt[b]{x^a}$

The underlying principle that we see from these rules is that multiplication of the base (x here) leads to addition in the exponents (a and b here), but multiplication in the exponents comes from nested exponentiation, for example, $(x^a)^b = x^{ab}$ from above. One point in this list is purely notational: Power(x, a) comes from the computer expression of mathematical notation.

A **logarithm** of (positive) x, for some ***base*** b, is the value of the exponent that gets b to x:

$$\log_b(x) = a \implies b^a = x.$$

A frequently used base is $b = 10$, which defines the **common log**. So, for example,

$$\log_{10}(100) = 2 \implies 10^2 = 100$$
$$\log_{10}(0.1) = -1 \implies 10^{-1} = 0.1$$
$$\log_{10}(15) = 1.176091 \implies 10^{1.1760913} = 15.$$

Another common base is $b = 2$:

$$\log_2(8) = 3 \implies 2^3 = 8$$
$$\log_2(1) = 0 \implies 2^0 = 1$$
$$\log_2(15) = 3.906891 \implies 2^{3.906891} = 15.$$

Actually, it is straightforward to change from one logarithmic base to another. Suppose we want to change from base b to a new base a. It turns out that we only need to divide the first expression by the log of the new base ***to the old base***:

$$\log_a(x) = \frac{\log_b(x)}{\log_b(a)}.$$

For example, start with $\log_2(64)$ and convert this to $\log_8(64)$. We simply have to divide by $\log_2(8)$:

$$\log_8(64) = \frac{\log_2(64)}{\log_2(8)}$$
$$2 = \frac{6}{3}.$$

We can now state some general properties for logarithms of all bases.

Basic Properties of Logarithms

$\rightarrowtail$	Zero/One	$\log_b(1) = 0$
$\rightarrowtail$	Multiplication	$\log(x \cdot y) = \log(x) + \log(y)$
$\rightarrowtail$	Division	$\log(x/y) = \log(x) - \log(y)$
$\rightarrowtail$	Exponentiation	$\log(x^y) = y \log(x)$
$\rightarrowtail$	Basis	$\log_b(b^x) = x$, and $b^{\log_b(x)} = x$

A third common base is perhaps the most interesting. The **natural log** is the log with the irrational base: $e = 2.718281828459045235\ldots$. This does not seem like the most logical number to form a useful base, but in fact it turns out to be so. This is an enormously important constant in our numbering system and appears to have been lurking in the history of mathematics for quite some time, however, without substantial recognition. Early work on logarithms in the seventeenth century by Napier, Oughtred, Saint-Vincent, and Huygens hinted at the importance of e, but it was not until Mercator published a table of "natural logarithms" in 1668 that e had an association. Finally, in 1761 e acquired its current name when Euler christened it as such.

Mercator appears not to have realized the theoretical importance of e, but soon thereafter Jacob Bernoulli helped in 1683. He was analyzing the (now-famous) formula for calculating compound interest, where the compounding is done continuously (rather than a set intervals):

$$f(p) = \left(1 + \frac{1}{p}\right)^p .$$

Bernoulli's question was, what happens to this function as p goes to infinity? The answer is not immediately obvious because the fraction inside goes to zero, implying that the component within the parenthesis goes to one and the exponentiation does not matter. But does the fraction go to zero faster than the exponentiation grows ever larger? Bernoulli made the surprising discovery that this function in the limit (i.e., as $p \to \infty$) must be between 2 and 3. Then what others missed Euler made concrete by showing that the limiting value of this function is actually e. In addition, he showed that the answer to Bernoulli's question could also be found by

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \dots$$

(sometimes given as $e = \frac{1}{1!} + \frac{2}{2!} + \frac{3}{3!} + \frac{4}{4!} + \dots$). Clearly this (Euler's expansion) is a series that adds declining values because the factorial in the denominator will grow much faster than the series of integers in the numerator.

Euler is also credited with being the first (that we know of) to show that e, like π, is an **irrational number**: There is no end to the series of nonrepeating numbers to the right of the decimal point. Irrational numbers have bothered mankind for much of their recognized existence and have even had negative connotations. One commonly told story holds that the Pythagoreans put one of their members to death after he publicized the existence of irrational numbers. The discovery of negative numbers must have also perturbed the Pythagoreans because they believe in the beauty and primacy of natural numbers (that the diagonal of a square with sides equal to one unit has length $\sqrt{2}$ and that caused them great consternation).

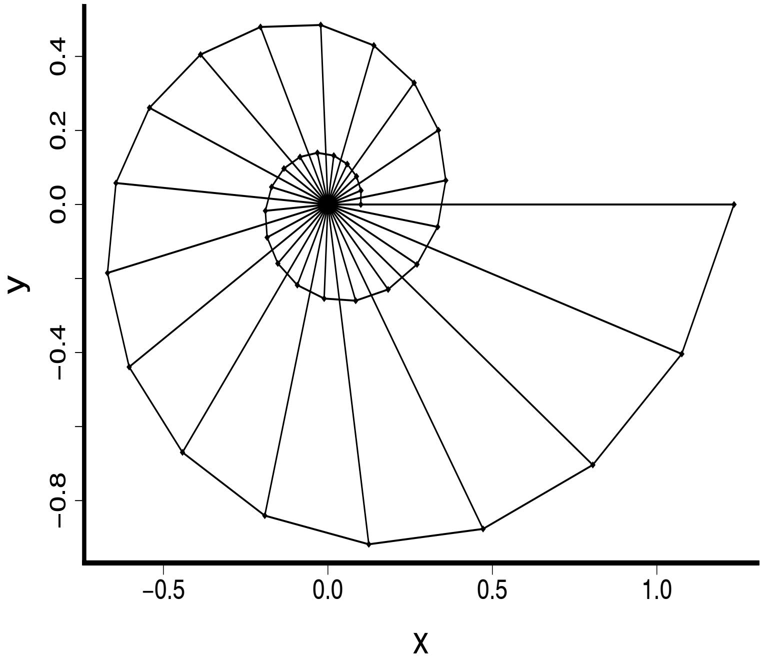
It turns out that nature has an affinity for e since it appears with great regularity among organic and physical phenomena. This makes its use as a base for the log function quite logical and supportable. As an example from biology, the chambered nautilus (***nautilus pompilius***) forms a shell that is characterized as "equiangular" because the angle from the source radiating outward is constant as the animal grows larger. Aristotle (and other ancients) noticed this as well as the fact that the three-dimensional space created by growing new chambers always has the same shape, growing only in magnitude. We can illustrate this with a cross section of the shell created by a growing spiral of consecutive right triangles (the real shell is curved on the outside) according to

$$x = r \times e^{k\theta} \cos(\theta) \qquad y = r \times e^{k\theta} \sin(\theta),$$

where r is the radius at a chosen point, k is a constant, θ is the angle at that point starting at the x-axis proceeding counterclockwise, and $\sin$, $\cos$ are functions that operate on angles and are described in the next chapter (see page 56). Notice the centrality of e here, almost implying that these mulluscs sit on the ocean floor pondering the mathematical constant as they produce shell chambers. A two-dimensional cross section is illustrated in Figure 1.7 ($k = 0.2$, going around two rotations), where the characteristic shape is obvious even with the triangular simplification.

Given the central importance of the natural exponent, it is not surprising that

Fig. 1.7. Nautilus Chambers



the associated logarithm has its own notation:

$$\log_e(x) = \ln(x) = a \implies e^a = x,$$

and by the definition of e

$$\ln(e^x) = x.$$

This inner function (e^x) has another common notational form, $\exp(x)$, which comes from expressing mathematical notation on a computer. There is another notational convention that causes some confusion. Quite frequently in the statistical literature authors will use the generic form $\log()$ to denote the natural logarithm based on e. Conversely, it is sometimes defaulted to $b = 10$ elsewhere (often engineering and therefore less relevant to the social sciences). Part of the reason for this shorthand for the natural log is the pervasiveness of e in the mathematical forms that statisticians care about, such as the form that defines the $normal$ probability distribution.

1.8 New Terminology

Exercises

1.1 Simplify the following expressions as much as possible:

$$(-x^4y^2)^2 \qquad 9(3^0) \qquad (2a^2)(4a^4)$$

$$\frac{x^4}{x^3} \qquad (-2)^{7-4} \qquad \left(\frac{1}{27b^3}\right)^{1/3}$$

$$y^7y^6y^5y^4 \qquad \frac{2a/7b}{11b/5a} \qquad (z^2)^4$$

1.2 Simplify the following expression:

$$(a+b)^2 + (a-b)^2 + 2(a+b)(a-b) - 3a^2$$

1.3 Solve:

$$\sqrt[3]{2^3} \qquad \sqrt[3]{27} \qquad \sqrt[4]{625}$$

1.4 The relationship between Fahrenheit and Centigrade can be expressed as $5f - 9c = 160$. Show that this is a linear function by putting it in $y = mx + b$ format with $c = y$. Graph the line indicating slope and intercept.

1.5 Another way to describe a line in Cartesian terms is the **point-slope form**: $(y - y') = m(x - x')$, where y' and x' are given values and m is the slope of the line. Show that this is equivalent to the form given by solving for the intercept.

1.6 Solve the following inequalities so that the variable is the only term on the left-hand side:

$$x - 3 < 2x + 15$$
$$11 - \frac{4}{3}t > 3$$
$$\frac{5}{6}y + 3(y-1) \leq \frac{11}{6}(1-y) + 2y$$

1.7 A very famous sequence of numbers is called the Fibonacci sequence, which starts with 0 and 1 and continues according to:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$$

Figure out the logic behind the sequence and write it as a function using subscripted values like x_j for the jth value in the sequence.

1.8 In the example on page 24, the cube law was algebraically rearranged to solve for S_A. Show these steps.

1.9 Which of the following functions are continuous? If not, where are the discontinuities?

$$f(x) = \frac{9x^3 - x}{(x-1)(x+1)} \qquad g(y,z) = \frac{6y^4z^3 + 3y^2z - 56}{12y^5 - 3zy + 18z}$$

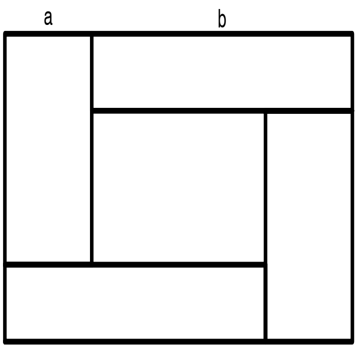
$$f(x) = e^{-x^2} \qquad f(y) = y^3 - y^2 + 1$$

$$h(x,y) = \frac{xy}{x+y} \qquad f(x) = \begin{cases} x^3 + 1\, x > 0 \\ \frac{1}{2}\; x = 0 \\ -x^2 x < 0 \end{cases}$$

1.10 Find the equation of the line that goes through the two points $\{[-1,-2],[3/2,5/2]\}$.

1.11 Use the diagram of the square to prove that $(a-b)^2 + 4ab = (a+b)^2$

(i.e., demonstrate this equality geometrically rather than algebraically with features of the square shown).



1.12 Suppose we are trying to put together a Congressional committee that has representation from four national regions. Potential members are drawn from a pool with 7 from the northeast, 6 from the south, 4 from the Midwest, and 6 from the far west. How many ways can you choose a committee that has 3 members from each region for a total of 12?

1.13 Sørensen's (1977) model of social mobility looks at the process of increasing attainment in the labor market as a function of time, personal qualities, and opportunities. Typical professional career paths follow a logarithmic-like curve with rapid initial advancement and tapering off progress later. Label y_t the attainment level at time period t and y_{t-1} the attainment in the previous period, both of which are defined over $\Re^+$. Sørensen stipulates:

$$y_t = \frac{r}{s}[\exp(st) - 1] + y_{t-1}\exp(st),$$

where $r \in \Re^+$ is the individual's resources and abilities and $s \in \Re^+$ is the structural impact (i.e., a measure of opportunities that become available). What is the domain of s, that is, what restrictions are necessary on what values it can take on in order for this model to make sense in that declining marginal manner?

1.14 The following data are U.S. Census Bureau estimates of population over a 5-year period.

Date	Total U.S. Population
July 1, 2004	293,655,404
July 1, 2003	290,788,976
July 1, 2002	287,941,220
July 1, 2001	285,102,075
July 1, 2000	282,192,162

Characterize the growth in terms of a parametric expression. Graphing may help.

1.15 Using the change of base formula for logarithms, change $\log_6(36)$ to $\log_3(36)$.

1.16 Glottochronology is the anthropological study of language change and evolution. One standard theory (Swadish 1950, 1952) holds that words endure in a language according to a "decay rate" that can be expressed as $y = c^{2t}$, where y is the proportion of words that are retained in a language, t is the time in 1000 years, and $c = 0.805$ is a constant. Reexpress the relation using "e" (i.e., 2.71...), as is done in some settings, according to $y = e^{-t/\tau}$, where τ is a constant you must specify. Van der Merwe (1966) claims that the Romance-Germanic-Slavic language split fits a curve with $\tau = 3.521$. Graph this curve and the curve from τ derived above with an x-axis along 0 to 7. What does this show?

1.17 Sociologists Holland and Leinhardt (1970) developed measures for models of structure in interpersonal relations using ranked clusters. This approach requires extensive use of factorials to express personal choices. The authors defined the notation $x^{(k)} = x(x-1)(x-2)\cdots(x-k+1)$. Show that $x^{(k)}$ is just $x!/(x-k)!$.

1.18 For the equation $y^3 = x^2 + 2$ there is only one solution where x and y are both positive integers. Find this solution. For the equation $y^3 = x^2 + 4$ there are only two solutions where x and y are both positive integers. Find them both.

1.19 Show that in general

$$\sum_{i=1}^{m} \prod_{j=1}^{n} x_i y_j \neq \prod_{j=1}^{n} \sum_{i=1}^{m} x_i y_j$$

and construct a special case where it is actually equal.

1.20 A **perfect number** is one that is the sum of its proper divisors. The first five are

$$6 = 1 + 2 + 3$$

$$28 = 1 + 2 + 4 + 7 + 14$$

$$496 = 1 + 2 + 4 + 8 + 16 + 31 + 62 + 124 + 248.$$

Show that 8128 are 33550336 perfect numbers. The Pythagoreans also defined **abundant numbers**: The number is less than the sum of its divisors, and **deficient numbers**: The number is greater than the sum of its divisors. Any divisor of a deficient number or perfect number turns out to be a deficient number itself. Show that this is true with 496. There is a function that relates perfect numbers to primes that comes from Euclid's ***Elements*** (around 300 BC). If $f(x) = 2^x - 1$ is a prime number, then $g(x) = 2^{x-1}(2^x - 1)$ is a perfect number. Find an x for the first three perfect numbers above.

1.21 Suppose we had a linear regression line relating the size of state-level unemployment percent on the x-axis and homicides per 100,000 of the state population on the y-axis, with slope $m = 2.41$ and intercept $b = 27$. What would be the expected effect of increasing unemployment by 5%?

1.22 Calculate the following:

$$\begin{aligned} 113 &\pmod 3 \\ 256 &\pmod{17} \\ 45 &\pmod 5 \\ 88 &\pmod{90}. \end{aligned}$$

1.23 Use Euler's expansion to calculate e with 10 terms. Compare this result to some definition of e that you find in a mathematics text. How accurate were you?

1.24 Use Stirling's Approximation to obtain 12312!. Show the steps.

1.25 Find the roots (solutions) to the following quadratic equations:

$$\begin{aligned} 4x^2 - 1 &= 17 \\ 9x^2 - 3x + 12 &= 0 \\ x^2 - 2x - 16 &= 0 \\ 6x^2 - 6x - 6 &= 0 \\ 5 + 11x &= -3x^2. \end{aligned}$$

1.26 The manner by which seats are allocated in the House of Representatives to the 50 states is somewhat more complicated than most people appreciate. The current system (since 1941) is based on the "method of equal proportions" and works as follows:

- Allocate one representative to each state regardless of population.
- Divide each state's population by a series of values given by the formula $\sqrt{i(i-1)}$ starting at $i = 2$, which looks like this for state j with population p_j:

$$\frac{p_j}{\sqrt{2 \times 1}}, \frac{p_j}{\sqrt{3 \times 2}}, \frac{p_j}{\sqrt{4 \times 3}}, \dots \frac{p_j}{\sqrt{n \times (n-1)}},$$

  where n is a large number.
- These values are sorted in descending order for all states and House seats are allocated in this order until 435 are assigned.

(a) The following are estimated state "populations" for the original 13 states in 1780 (Bureau of the Census estimates; the first official U.S. census was performed later in 1790):

Virginia	538,004
Massachusetts	268,627
Pennsylvania	327,305
North Carolina	270,133
New York	210,541
Maryland	245,474
Connecticut	206,701
South Carolina	180,000
New Jersey	139,627
New Hampshire	87,802
Georgia	56,071
Rhode Island	52,946
Delaware	45,385

Calculate under this plan the apportionment for the first House of Representatives that met in 1789, which had 65 members.

(b) The first apportionment plan was authored by Alexander Hamilton and uses only the proportional value and rounds down to get full persons (it ignores the remainders from fractions), and any remaining seats are allocated by the size of the remainders to give $(10, 8, 8, 5, 6, 6, 5, 5, 4, 3, 3, 1, 1)$ in the order above. Relatively speaking, does the Hamilton plan favor or hurt large states? Make a graph of the differences.

(c) Show by way of a graph the increasing proportion of House representation that a single state obtains as it grows from the smallest to the largest in relative population.

1.27 The Nachmias–Rosenbloom Measure of Variation (MV) indicates how many heterogeneous intergroup relationships are evident from the full set of those mathematically possible given the population. Specifically it is described in terms of the "frequency" (their original language) of observed subgroups in the full group of interest. Call f_i the frequency or proportion of the ith subgroup and n the number of these groups. The index is created by

$$MV = \frac{\text{“each frequency} \times \text{all others, summed”}}{\text{“number of combinations”} \times \text{“mean frequency squared”}}$$

$$= \frac{\sum_{i=1}^{n}(f_i \neq f_j)f_i f_j}{\frac{n(n-1)}{2}\bar{f}^2}.$$

Nachmias and Rosenbloom (1973) use this measure to make claims about how integrated U.S. federal agencies are with regard to race. For a population of 24 individuals:

(a) What mixture of two groups (say blacks and whites) gives the maximum possible MV? Calculate this value.

(b) What mixture of two groups (say blacks and whites) gives the minimum possible MV but still has both groups represented? Calculate this value as well.

1.9 Chapter Appendix: It's All Greek to Me

The following table lists the Greek characters encountered in standard mathematical language along with a very short description of the standard way that each is considered in the social sciences (omicron is not used).

Name	Lowercase	Capitalized	Typical Usage
alpha	α	–	general unknown value
beta	β	–	general unknown value
gamma	γ	Γ	small case a general unknown value, capitalized version denotes a special counting function
delta	δ	Δ	often used to denote a difference
epsilon	ϵ	–	usually denotes a very small number or error
zeta	ζ	–	general unknown value
eta	η	–	general unknown value
theta	θ	Θ	general unknown value, often used for radians
iota	ι	–	rarely used
kappa	κ	–	general unknown value
lambda	λ	Λ	general unknown value, used for eigenvalues
mu	μ	–	general unknown value, denotes a mean in statistics

Name	Lowercase	Capitalized	Typical Usage
nu	ν	–	general unknown value
xi	ξ	Ξ	general unknown value
pi	π	Π	small case can be: 3.14159..., general unknown value, a probability function; capitalized version should not be confused with product notation
rho	ρ	–	general unknown value, simple correlation, or autocorrelation in time-series statistics
sigma	σ	Σ	small case can be unknown value or a variance (when squared), capitalized version should not be confused with summation notation
tau	τ	–	general unknown value
upsilon	υ	Υ	general unknown value
phi	ϕ	Φ	general unknown value, sometimes denotes the two expressions of the normal distribution
chi	χ	–	general unknown value, sometimes denotes the chi-square distribution (when squared)
psi	ψ	Ψ	general unknown value
omega	ω	Ω	general unknown value