Essential Mathematics for the Political and Social Research

JEFF GILL
Cambridge University Press

Lecture Slides, Chapter 1: The Basics

Chapter 1 Objectives

- Give an introduction to practical mathematical and arithmetic principles.
- Review some notation and some symbology.
- Consider some abstract but basic mathematical foundations.
- Make sure that we're all at the same level before moving on to matrix algebra and calculus.

Why is mathematics important to social scientists?

- It allows us to communicate with each other in an orderly and systematic way.
- Ideas expressed mathematically can be more carefully defined and more directly communicated than with narrative language.
- 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.

What about social behavior in particular?

- Such phenomena are easy to observe but generally difficult to describe in simple mathematical terms.
- Substantial progress dates back only to the 1870s, starting with economics, and followed closely by psychology.
- 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,
  - the size of legislatures affects the length of decisions,

Essential Arithmetic Principles

- The order of operations matters:
  - perform operations on individual values first,
  - evaluate parenthetical operations next,
  - do multiplications and divisions in order from left to right,
  - do additions and subtractions from left to right.
- Example

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

Zero

- Zero represents a special number in mathematics.
- The history of mathematics is very much tied to the interpretation of zero and infinity.
- Multiplying by zero produces zero and adding zero to some value leaves it unchanged.
- The worry about with zero is that dividing any number by zero ($x/0$ for any x) is *undefined*.
- $0 \times \infty$ is also undefined because it would otherwise lead to series pathologies, such as:

$$\begin{aligned} 0 \times \infty &= k \\ \infty &= k/0 \end{aligned}$$

Some Basic Functions

- 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 *radical index*
- For example,
$$\sqrt[3]{8} = 8^{\frac{1}{3}} = 2$$
because $2^3 = 8$.

Example: Explaining Why People Vote

- Anthony Downs (1957) claimed that a rational voter who values her time and resources weighs the cost of voting against the gains received from voting in terms of "utility."

- We should make this idea more "mathematical" by articulating 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 $R = PB - C$.

- Now consider these statements:

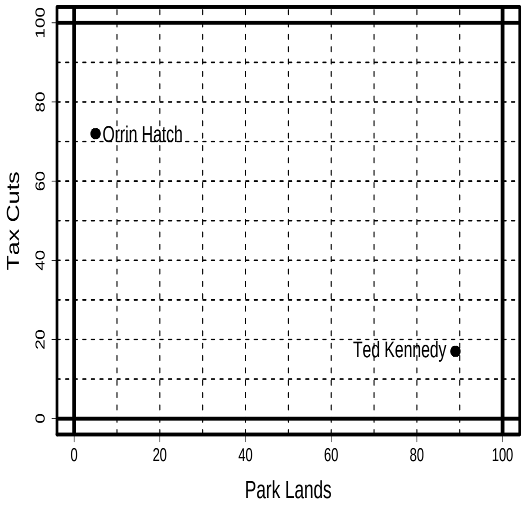
  - The voter will abstain if $R < 0$.
  - The voter may still not vote even if $R > 0$ if there exist other competing activities that produce a higher R.
  - If P is very small (i.e., it is a large election with many voters), then it is unlikely that this individual will vote (*the paradox of participation*)..

Some Basic Terms

- A *variable* is just a *symbol* that represents a single number or group of numbers, X, Y, a, b, and so on.
- Unless otherwise stated, variables are assumed to be defined on the *Cartesian coordinate system*:
  - given two variables x and y,
  - there is an assumed perpendicular set of axes where the x-axis (always given horizontally) is crossed with the y-axis (usually given vertically) that range from $-\infty$ to $+\infty$,
  - such that the number pair (x, y) defines a *point* on the two-dimensional graph,
  - and there is actually no restriction to just two dimensions.
- 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*.

Example: 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*
- The figure shows approximate ideal points for a very liberal and very conservative senator on two proposed projects.



- 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 $x = 89, y = 17$.

Example: Graphing Ideal Points in the Senate

- Consider a hypothetical trade-off between two bills competing for limited federal resources: 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.
- The point of this 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.

The Real Line

- The basis for the Cartesian coordinate system is that the measurement of any dimension on the *real line*:
  - ranges from negative infinity to positive infinity (although we obviously cannot *draw* it that way),
  - labeled $\Re$
  - 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 ($\sqrt{2} = 1.41421356237309504880168872420969807856967187537694807317\ldots$).

Sets Along the Real Line

- These sets are a bounded range of real numbers, not necessarily contiguous.
- 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 $0 \le \delta \le 1$) 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.
  - A *nonconvex set* example is: $\{[2:4],[5:6]\}$.

Interval Notation

- An open interval excludes the end-point denoted with parenthetical forms "(" and ")".
- A closed interval denoted with bracket forms "[" and "]".
- the curved forms "{" and "}" are usually reserved for set notation.
- Variations:

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

- Sometimes you see *comma notation* instead of *colon notation*: $[0, 100]$.

Indexing and Referencing

- A common notation is indexing observations on some variable by the use of *subscripts*, $\mathbf{X} = \{X_1, X_2, X_3, \ldots, X_{433}, X_{434}, X_{435}\}$.
- This is a lot cleaner and more mathematically useful than creating many variables with different names, especially when they are related cases.
- To calculate the mean (average), 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*.

- The *summation operator* 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:

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

where we say that i *indexes* X in the summation, and $n = 435$ is the stopping point.

Product Operator

- A related notation is the *product operator*: 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).

- Also uses 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.

- Suppose $p_1 = 0.2$, $p_2 = 0.7$, $p_3 = 0.99$, $p_4 = 0.99$, $p_5 = 0.99$. Then

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

More On Subscripts

- 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:
  - X^2 is the square of X,
  - but $X^{[2]}$ and $X^{(2)}$ are indexed values.
- Superscripts are very helpful with *stochastic processes*, including *Markov chains*.
- Another notation comes from computer usage: $X1, X2, \ldots$.

Specific Mathematical Use of Terms

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 (expressible in fractional form)
$\mathbf{Q}^+$	the set of positive rational numbers
$\mathbf{Q}^-$	the set of negative rational numbers
$\mathcal{C}$	the set of complex numbers (those based on $\sqrt{-1}$).

Logical Modifiers (negated with /)

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

Examples

- $3 - 4 \in \mathfrak{R}^- \because 3 < 4$.
- $2 \in \mathcal{I}^+ \therefore 2 \in \mathfrak{R}^+$.
- $x \in \mathcal{I}$ and $x \neq 0 \Longrightarrow x \in \mathcal{I}^-$ or $\mathcal{I}^+$.
- "If x is a nonzero integer, it is either a positive integer or a negative integer.
- $\forall x \in \mathcal{I}^+$ and $x \neg$prime, $\exists y \in \mathcal{I}^+ \ni x/y \in \mathcal{I}^+$.
- "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."

Set Theory Notation

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

Example: $A \subset B$ for $A = \{2, 4\}$, $B = \{2, 4, 7\}$.

Symbols Restricted to Numerical Values

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: $X\vert Y=3$

Symbols Related to Extremes

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

Equations

- 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.
- Generalizing with other relations:

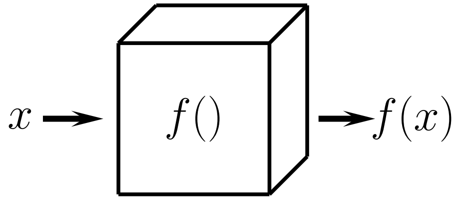
Symbol	Meaning
$<$	less than
$\leq$	less than or equal to
$\ll$	much less than
$>$	greater than
$\geq$	greater than or equal to
$\gg$	much greater than
$\approx$	approximately the same
$\cong$	approximately equal to
$\lessapprox$	approximately less than (also $\lesssim$)
$\gtrapprox$	approximately greater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

Functions

- A *mathematical function* is a "mapping" (i.e., specific directions), which gives a correspondence from one measure onto exactly one other for that value.
- 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., $f(x) = \text{round}(x)$).

Functions

- One way of thinking about functions is that they are a "machine" for transforming values, sort of a box as in the figure:



- Visualize values, x, going in and some modification of these values, $f(x)$, coming out where the instructions for this process are contained in the "recipe" given by $f()$.

- 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 $u()$.

Example of Functions

- Consider the following function operating on the variable x:

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

  a mapping from x to $f(x)$ is the process that squares x and subtracts 1.

- Listing a set of inputs, we can define the corresponding set of outputs:

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

Additional Functional Notation

- Necessary 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.

Additional Functional Notation, Examples

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

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

Properties. . .

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

- ⤚ A function is *continuous* if it has no “gaps” in its mapping from x to y.
- ⤚ A function is *invertible* if its reverse operation exists: $g^{-1}(y) = x$, where $g^{-1}(g(x)) = x$.

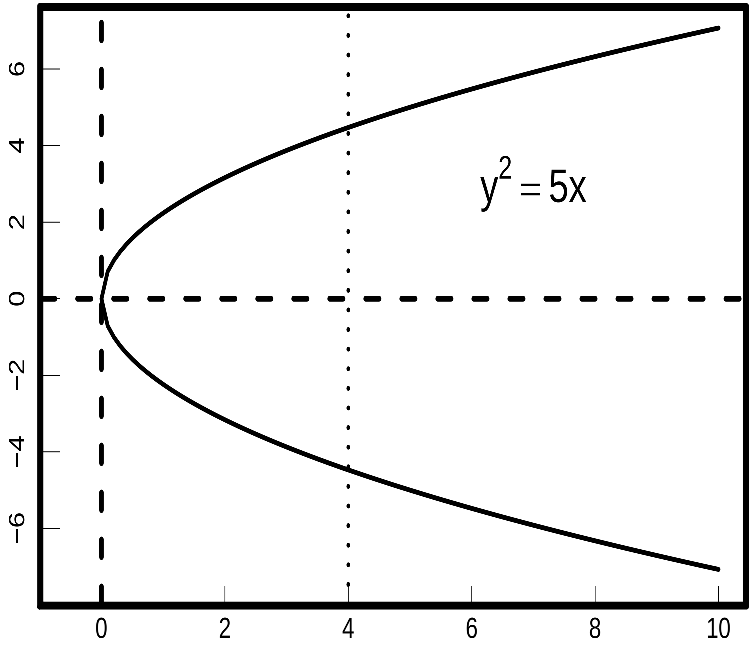
- Non-continuous example: $f(x) = \frac{1}{x-1}$ is not continuous at 1.
- Non-invertible example: $f(x) = x^2$, because there are two possible answers.

Relations

- 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.
- 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.
- We can also see this graphically in the following figure 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).

Relations

Figure 1: A Relation That Is Not a Function



Functions as Mappings

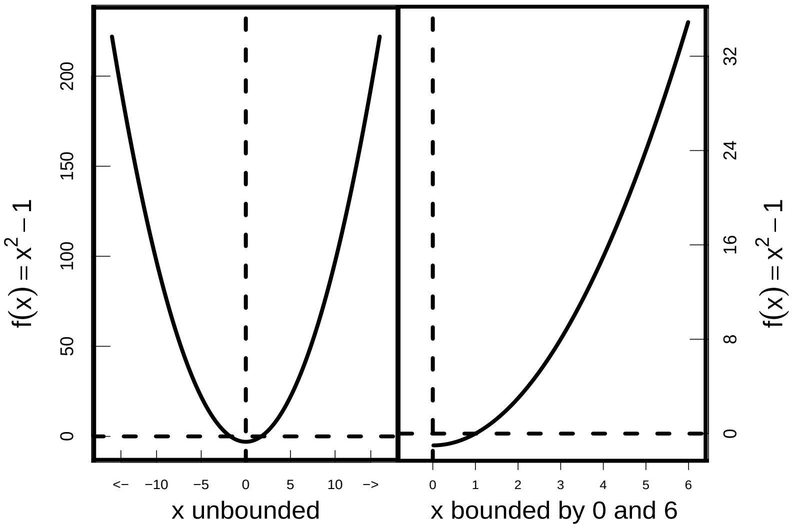
- Often a function is explicitly defined as a mapping between elements of an *ordered pair* (x, y).
- The function $f(x) = y$ maps the ordered pair x, y such that for each value of x there is exactly one y (the order of x before y matters).
- For example, the following set of ordered pairs:

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

  can be mapped by the function: $f(x) = x^2 - 3$.
- 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
- Two examples are given in the following figure, which is drawn on the (now) familiar Cartesian coordinate system.

Functions as Mappings

Figure 2: Relating x and $f(x)$



Odd and Even Functions

- A function can also be even or odd, defined by

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

- 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.
- But $f(x) = x^3$ is odd because the negative sign perseveres for a negative x.
- Regretfully, functions can also be neither even nor odd without domain restrictions.

Linear Functions

- 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.

Legislative Example

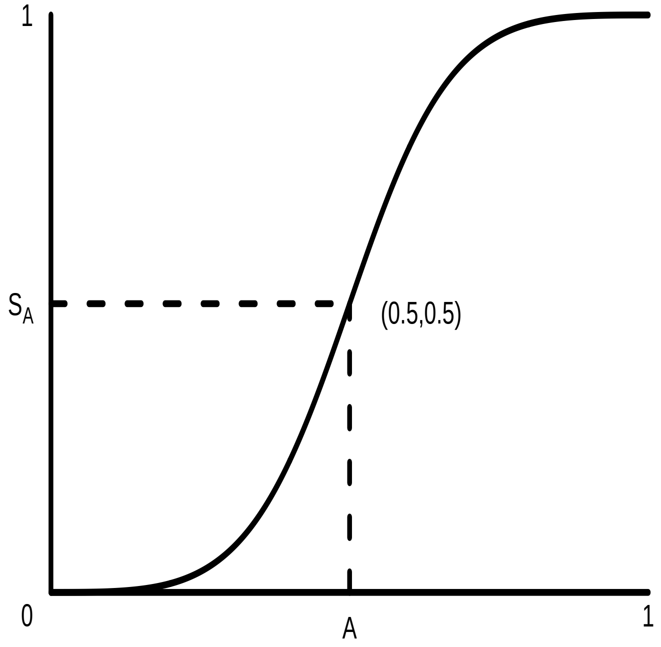
- A standard, though somewhat maligned, theory from the study of elections is due to Parker's (1909) empirical research in Britain, which was later popularized in that country by Kendall and Stuart (1950, 1952).
- He looked at systems with two major parties whereby the largest vote-getter in a district 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.
- 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.
- 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.

Legislative Example

- Define x as the ratio of votes for the party with proportion A over the party with proportion B.
- Expressing the cube law in this notation yields
$$f(x) = x^3$$
for the function determining seats.
- 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 $B = 1 - A$.
- 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, which produces
$$S_A = \frac{A^3}{1 - 3A + 3A^2}.$$

Legislative Example

- 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 the figure.



Applying Functions: The Equation of a Line

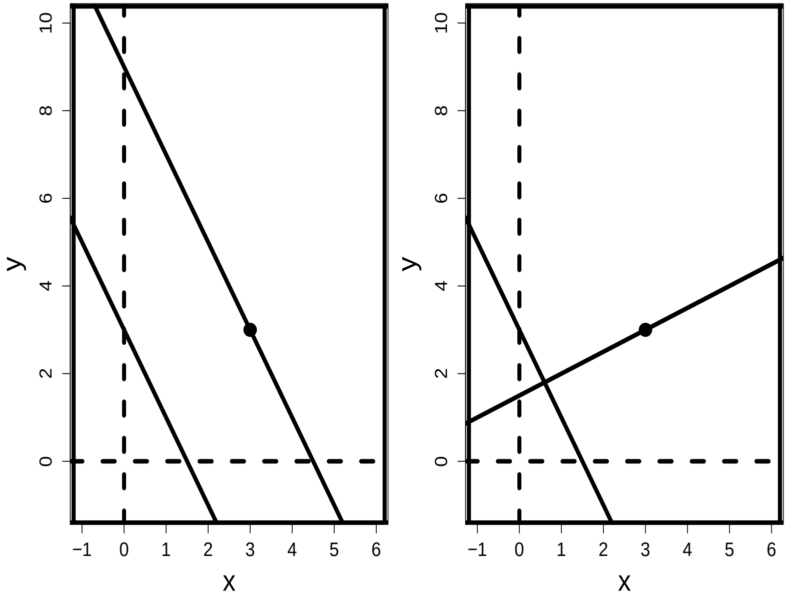
- Recall the familiar expression of a line in Cartesian coordinates usually given as $y = mx + b$, where m is the slope of the line (the change in y for a one-unit change in x) and b is the point where the line intercepts the y-axis.
- Clearly this is a (linear) function in the sense described above 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.

Applying Functions: The Equation of a Line

- We want to find the equation of the line that goes through the two points $\{[2,1],[3,5]\}$.
- For one unit of increasing x we get four units of increasing y.
- Since slope is "rise over run," then:
$$m = \frac{5-1}{3-2} = 4.$$
- To get the intercept 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:
$$\begin{aligned} y &= mx + b \\ 1 &= 4(2) + b \\ b &= 1 - 8 = -7. \end{aligned}$$

Applying Functions: The Equation of a Line

Figure 3: Parallel and Perpendicular Lines



Parallel Lines

- 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 the following figure:

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

- The parallel line is given by $L_2 : y = -2x + 9$.

Parallel Lines

- 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:

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

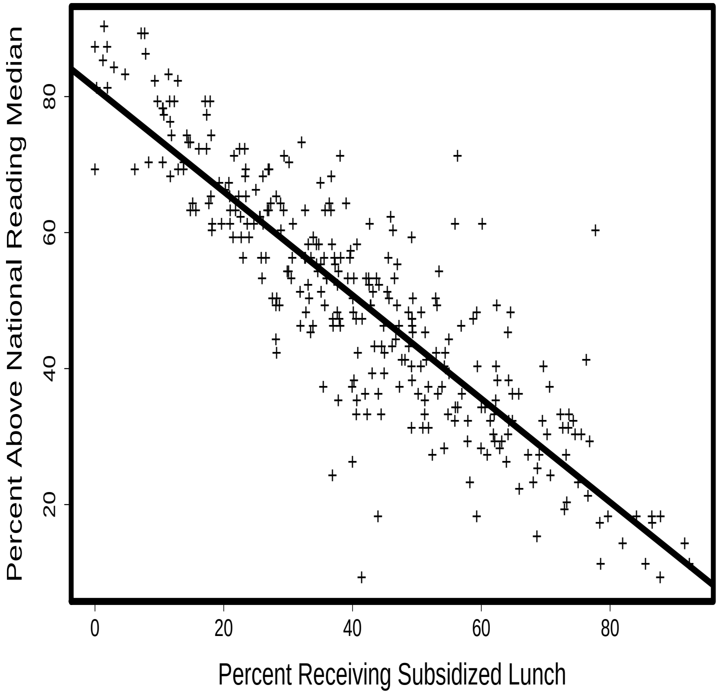
- This gives us $L_2 \perp L_1$, where $L_2 : y = \frac{1}{2}x + \frac{3}{2}$.

Example: Child Poverty and Reading Scores

- Despite overall national wealth, a surprising number of U.S. school children live in poverty.
- 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 relationship is graphed 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*.

Example: Child Poverty and Reading Scores

Figure 4: Poverty and Reading Test Scores



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,

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

- This function grows rapidly for increasing values of x, and sometimes the result overwhelms commonly used hand calculators.

The Factorial Function

- 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).

- Sometimes this problem can be avoidied with *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 $e \approx 2.71$.

Example: 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?
- 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.

Example: Coalition Cabinet Formation

- 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):

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

- 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.

The Modulo Function

- A function that has special notation is the *modulo function*, which deals with the *remainder* from a division operation.
- First 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).
- 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.
- 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).

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.

- The convention is that a polynomial degree (power) is designated by its largest exponent with regard to the variable.

Roots of Polynomial Functions

- 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: the only root of this function is at the point $x = 0$.

Quadratic Polynomials

- In many 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

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

- Sometimes they require 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 $f(x) = x^2 - 4x - 5 \equiv 0$.

Logarithms and Exponents

- Exponents and logarithms ("logs" for short) confuse many people.
- Basically exponents make convenient the idea of multiplying a number by itself (possibly) many times, and a logarithm is just the opposite operation.
- Rules for exponents:

Key Properties of Powers and Exponents

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

Logarithms and Exponents

- 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.

- 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,

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

Another common base is $b = 2$:

$$\begin{aligned} \log_2(8) = 3 &\quad\implies 2^3 = 8 \\ \log_2(1) = 0 &\quad\implies 2^0 = 1 \\ \log_2(15) = 3.906891 &\quad\implies 2^{3.906891} = 15. \end{aligned}$$

Logarithms and Exponents

- It is straightforward to change from one logarithmic base to another.
- Suppose we want to change from base b to a new base a.
- 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 have to divide by $\log_2(8)$:

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

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 $b^{\log_b(x)} = x$

The Natural Log

- The *natural log* is the log with the irrational base: $e = 2.718281828459045235\ldots$.
- 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.
- Jacob Bernoulli in 1683 was analyzing the (now-famous) formula for calculating compound interest, where the compounding is done continuously (rather than at 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?
- Bernoulli made the surprising discovery that this function in the limit (i.e., as $p \to \infty$) must be between 2 and 3

The Natural Log

- 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
$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots$$
- Euler is also credited with being the first (that we know of) to show that e, like π, is an *irrational number*.
- Nature has an affinity for e since it appears with great regularity among organic and physical phenomena.

Notation for Natural Log

- General notation:

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

- By the definition of e

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

- There is an alternative notation, $\exp(x)$, which comes from expressing mathematical notation on a computer.