Fun with Functions

Paul E. Johnson\textsuperscript{1} \textsuperscript{2}

\textsuperscript{1}Department of Political Science

\textsuperscript{2}Center for Research Methods and Data Analysis, University of Kansas

2015
Outline

1. Linear relationships
2. Nonlinear Overview
3. Splines and Segments
4. Polynomials
5. Exponentials
6. Logarithms
Ground Rules for this Lecture

- There’s no substitute for an actual math class in which you do homework and take tests
- What can this lecture provide?
  - reminders of some important terms that would probably be covered in a math class called “pre-calculus”
  - hints about which particular things might be most important in graduate research, so you can refresh your understanding (we see logarithms much more often than sine and cosine, for example)
  - Something you can look up in future to get a quick reminder of same 😊
Outline

1. Linear relationships
2. Nonlinear Overview
3. Splines and Segments
4. Polynomials
5. Exponentials
6. Logarithms
Imagine a relationship between 2 numeric variables

- Don’t be distracted by statistics. Don’t worry about data. This is your (last) chance to theorize freely
- Try to imagine two things that are related, possibly even “causally related”
- Try to make a statement like
  
  “If “?” goes up, then I expect “?” goes up as well.

  For example,
  
  - as a student’s investment of time in studying goes up, that student’s grades go up
  - as a car’s price goes up, its driver is more satisfied (and less likeable 😊)
A Linear relationship is the simplest model

- Common (not required) to name variables $x$ and $y$
  - $x$ is a predictor
  - $y$ is an outcome
  
  \[ y = \beta_0 + \beta_1 x \]  

- The outcome is equal to a fixed amount ($\beta_0$) plus a proportion of $x$
  (literally, $\beta_1 x$)

- $\beta_0$, $\beta_1$, are often called:
  - parameters, or
  - coefficients
  - weights

- $\beta_0$ sometimes called the “constant”
  or “y-intercept”

- $\beta_1$ is the “slope coefficient”
Parts of a Linear equation

Example:
\[ y = 3 + 1.3 \cdot x \]

- \( \beta_0 = 3 \). The “constant”.
  - When \( x_i = 0 \), the outcome will be 3.
- \( \beta_1 = 1.3 \). The slope.
  - The slope: 1.3 is the “rise over run”
  - For each 1 unit increase in \( x_i \), the outcome increases by 1.3.
Labels will vary across books

- In junior high school, many of us are taught the formula

  \[ y = m \cdot x + b \]

  - The slope is \( m \) and the intercept is \( b \).

- It is much more common in research to use one letter for all parameters in a line

  - Any symbol can be used, not always \( \beta \)

  \[ y = \xi_0 + \xi_1 x \]

- Sometimes I use letters to feel less self-conscious (I’m not Greek, after all)

  \[ y = b_0 + b_1 x \]
Use any names you want

- Very common to say that $x$ is the predictor variable and $y$ is the output, but that’s not required.
- Perhaps you want something that is easier to remember:

\[
\text{fahrenheit} = \beta_0 + \beta_1 \text{celsius}, \quad \beta_0 = 32, \quad \beta_1 = \frac{9}{5}
\]

\[
\text{mpg} = \beta_0 + \beta_1 \text{weight}
\]

- To convey the idea that we want to “plug in” several different values of $x$ to get several different values of $y$, we often write with a subscript $i$.

\[
y_i = \beta_0 + \beta_1 x_i
\]

- This makes it more clear that $\beta_0$ and $\beta_1$ are fixed values, that apply to different instances of $x_i$. 

Doman, Range, Real Number Line

- The **domain** is the set of values that $x$ might assume.
- The **range** is the set of values that $y$ might assume.
- Since a linear equation can “go on forever”, then we usually assume the domain is the real number line, $\mathbb{R}$. And the range is too.
- But you are free to restrict that, just by saying so. The following line applies for $x$ less than 40.

$$y = 0.29 + 1.02x, \ x < 40$$
The Fitted "regression" Line

- Data analysis often presents “noisy data”
- The predicted value of the outcome is a formula, say

\[
\text{income} = -2853.6 + 899.8 \text{ education}
\]

- A 1 unit increase in education “is associated with” (causes?) a 898.8 increase in income
- Don’t worry that the observed points do not lie exactly “on the line” at the moment.

**Question:** Why does it appear the y-intercept is 3000, rather than the correct $-2853.6$
Try this in R

We don’t plan to teach you any R code, but if you have R on your computer, you can type in a few commands to figure this out for yourself. Here, we use R as a simple “graphing calculator”

\[
\text{curve}(0.1 + 0.2 \times x, \text{ from} = 0, \text{ to} = 4.1)
\]

- You can replace the slope and intercept, as well as the values of from and to.

\[
\text{curve}(0.1 + 0.2 \times x, \text{ from} = 3, \text{ to} = 9.1, \text{ ylim} = \text{ c }(-4, 5))
\]

- As soon as you make that work, we’ll try some more.

- Almost forgot to mention: If you make a typographical error, hit the “UP” arrow key. It will give back your mistake so you can edit it.
Draw 2 lines, tell a story about why you might need 2 lines

\begin{verbatim}
curve(0.1 + 0.2 * x, from = 0.5, to = 4.1, xlim = c(-1, 5), ylim = c(-1, 1))
curve(0.3 + 0.25 * x, from = 0.5, to = 4.1, add = TRUE)
\end{verbatim}

- You need to run those 2 commands, in order.
- Fiddle the coefficients.
  - First, adjust the formulae, the first part $0.1 + 0.2 \times x$
  - The other elements in the commands are just there to control the range of calculation and the plot size.
My Story About 2 Lines

- One line for people from the USA
Draw several lines

- One line for people from the USA
- One line for people from Canada
Still meaningful if $x$ is equal to 0 or 1

- A variable that is coded 0 or 1 indicates
- the absence or presence of a quality
- a categorical difference, “east, west” or “male, female”
- A formula

$$y = \beta_0 + \beta_1 x$$

is still meaningful
To dramatize it, people will draw it like steps.

\[ y = -0.7 + 1.95 \times x, \quad x \in \{0, 1\} \]
Outline

1. Linear relationships
2. Nonlinear Overview
3. Splines and Segments
4. Polynomials
5. Exponentials
6. Logarithms
What if the relationship is very wiggly?

If the relationship is like this

we are in for an unhappy career if we insist on the simple linear model.
Many “true relationships” are not linear

- **Galileo (@1590)**
  - \( d = \frac{1}{2}gt^2 \) (distance fallen is time squared times a constant)
  - \( t = \sqrt{\frac{2d}{g}} \) (time to fall distance \( d \), arbitrary constant \( g \))

- **Einstein (1905)** \( E = mc^2 \) (energy is mass times speed of light squared)
Lacking a good reason to suppose a relationship is nonlinear, we start with the linear hypothesis.

Why?

- The straight line is the simplest relationship possible.
- Ockham’s razor (William of Ockham, 1287-1347), a somewhat vague, often cited heuristic:

  *If presented with competing theories that fit the facts equally well, the simpler one is to be preferred*

- Any “wiggly” model can be added on to the linear model if needed.
The “True” Relationship is Approximately Linear

- The relationship is almost linear, so approximate it.
- Predictions won’t be much different
- Slope of line in that region won’t be much different
The “True” Model is Approximately Linear

- If data is collected, and it contains any “measurement error” at all, then a “line” and a “gradual curve” will not be distinguishable.
The Straight Line Model Is Not Useful If...

- Suppose the true “full relationship” very nonlinear
- Obviously, the “straight line theory” is wrong here

\[ y_i = 3 + 13.4x_i - 0.15x_i^2 \]
We might approximate on a Narrower Range

- On a small section of $x$, we can get pretty close.

![Graph showing linear and nonlinear curves](image-url)
People often emphasize Taylor’s Theorem

- From the first semester of calculus, there is a result that states that any function $f(x)$, can be approximated by an additive weighting of $x$, $x^2$, $x^3$, ... and some carefully chosen coefficients.

- The first order linear approximation of $f(x)$

  $$y = \beta_0 + \beta_1 (x - x_0)$$

- If you don’t project too far from $x_0$, a linear approximation stays close to $f(x)$
Approximating from a bad spot? Not Awesome
People often emphasize Taylor’s Theorem

- If the prediction is poor, however, Taylor’s Theorem holds that the additive approximation can be made “arbitrarily close” to $f(x)$ by adding more terms.

$$y = \beta_0 + \beta_1(x - x_0) + \beta_2(x - x_0)^2 + \beta_3(x - x_0)^3 + \text{possibly more terms}$$

- This is important in theoretical derivations, perhaps less so in “real life” function finding because it does not promise us that the powered terms will follow any particular pattern.
Outline

1. Linear relationships
2. Nonlinear Overview
3. Splines and Segments
4. Polynomials
5. Exponentials
6. Logarithms
Consider A Series of Segments

- A straight line will approximate any small part of the big parabola.
A Linear Spline Model

- Choose “knot” points, \( k_1, k_2, k_3 \)
- “Snap” the line at those points
- G.E.P. Box, “Essentially, all models are wrong, but some are useful.”
A Linear Spline Model

- Can think of that as 3 separate equations, one for each line segment (between knots)
- Or we can write one equation that is smart enough to track the changes.
- This uses a couple of handy functions, the “plus” and the “bump” functions.
Writing out the equations for a spline

- In order to write this down, we can use a special “plus” formula that equals 0 when $x$ is smaller than the knot and then it is $(x - k)$ above

$$\text{plus}(x, k) = \begin{cases} 
0 & \text{if } x < k \\
 x - k & \text{if } x \geq k 
\end{cases}$$

- Note the linear spline results from “piling” a series of plus functions together

Before $k_1$:
$$y = \beta_0 + \beta_1 x$$

After $k_1$, before $k_2$:
$$y = \beta_0 + \beta_1 x + \beta_2 \text{plus}(x, k_1)$$

After $k_2$:
$$y = \beta_0 + \beta_1 x + \beta_2 \text{plus}(x, k_1) + \beta_3 \text{plus}(x, k_2)$$

- The $\beta_2$ and $\beta_3$ are “slope shifts”: the proportional effect of $x_i$ “jumps”
Why bother with that “plus” stuff?

- If you write separate equations, and you are not very careful, you end up with gaps.
- Some theories might call for the gaps—"break points"—but theories are often based on the idea there is an unbroken curve we want to understand.
- Plus function: convenient way to have bends, not breaks, at the knots.
The plus saves you from having to figure out these separate intercepts

- Note each separate line has its own slope and intercept
- The plus approach gives that part automatically
If You Want Break Points

- Introduce a “bump” function that is 0 for $x$ below a knot point and 1 after.

\[
bump(x, k) = \begin{cases} 
0 & \text{if } x < k \\
1 & \text{if } x \geq k
\end{cases}
\]

- Then we include both a bump and a plus function to represent a slope change at a point where the relationship is not continuous.

Before $k_1$:

\[
y = \beta_0 + \beta_1 x
\]

After $k_1$:

\[
\beta_0 + \beta_1 x + \beta_2 plus(x, k_1) + \beta_3 bump(x, k_1)
\]

- The $\beta_2$ is a slope change
- $\beta_3$ is the “gap” in the relationship, the magnitude of the break
The size of the gap depends on where you measure!

- Measure the gap when
  - when $x$ is 0? Or
  - when $x$ is $k_1$?
- You can exaggerate the size of the gap by measuring at $x = 0$.
- Perhaps $x = 0$ is the important gap, perhaps not. Can you make an argument for either side?
Either way, we have 2 lines with different equations

- The Gap is, of course, bigger when \( x = 0 \). Impress your friends!
- If we are looking for a time-based change at time \( k_1 \), perhaps the gap worth discussing is at \( x = k_1 \)
- If we use the plus and bump functions, the coefficient represents the gap at \( x \)

- If we simply write separate lines, we end up thinking of the gap as the difference when \( x \) is 0, i.e, \( |\beta_0 - \beta_2| \) in the following

\[
y = \begin{cases} 
\beta_0 + \beta_1 x, & x < k_1 \\
\beta_2 + \beta_3 x, & x \geq k_1 
\end{cases}
\]
Can have “curves” between knots

- Until now, we used straight lines to connect the knot points.
- Possible to have *wiggly* shaped connectors between knots.
- In order to do that, we have to talk about ways to get wiggles.
- In more advanced context, the transformed predictor columns are called “basis functions” and we end up with a great deal of variety. Natural cubic: only cubic terms between knots!
Outline

1. Linear relationships
2. Nonlinear Overview
3. Splines and Segments
4. Polynomials
5. Exponentials
6. Logarithms
$x^2$ is U-shaped

Try this in R

```r
curve(x^2, from = -5, to = 5,
     xlim = c(-5, 15), ylim = c(-15, 15))
curve((x-4)^2, add = TRUE)
curve((x-10)^2, add = TRUE)
curve(-(x-10)^2, add = TRUE, col = "red")
```
x to the power of 2

Did you end up with this?
Flip!

$y = 0.2 \cdot x^2$

$y = -0.2 \cdot x^2$
Include $x$ and powers of $x$ as predictors

- It's called a polynomial if it has several powers added in:
  \[ y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots \]

- It's called a "quadratic equation" if it has no higher power than 2.
  \[ y = \beta_0 + \beta_1 x + \beta_2 x^2 \]
Functions

Polynomials

Quadratic

- By adjusting the coefficients \( \beta_0 + \beta_1 x + \beta_2 x^2 \), then, we can make the plot
  - either a “hill” or a “valley”
  - shift it side to side
  - or up and down

- Try this, you’ll see what I mean

```r
curve(1 + 2*x + 0.4*x^2, from = -20, to = 30)
curve(100 + 2*x + 0.4*x^2, from = -20, to = 30, add = TRUE, col = "red")
curve(100 + 6*x + 0.4*x^2, from = -20, to = 30, add = TRUE, col = "green")
abline(v=0, col = "gray70")
```
In college, I learned

- If \( \beta_2 < 0 \) then this is a “hill” shaped function
- \( \beta_2 > 0 \), then this is a “U” shaped function
- The “peak” or “bottom” occurs where

\[
x = \frac{-\beta_1}{2\beta_2}
\]
In high school, they used different letters

\[ y = ax^2 + bx + c \]

And they made us memorize

\[ -b \pm \sqrt{b^2 - 4ac} \]
\[ 2a \]

I never knew of a reason why anybody wanted those values, but I memorized the formula anyway 😊
Substantive Meaning of Quadratic

- Rewrite this

\[ y = \beta_0 + \beta_1 x + \beta_2 x^2 \]  \hspace{1cm} (2)

as

\[ y = \beta_0 + (\beta_1 + \beta_2 x) x \]  \hspace{1cm} (3)

- We assert the slope of the relationship is \( \beta_1 + \beta_2 x \).
- If \( \beta_2 > 0 \), then the effect of \( x \) gets bigger as \( x \) gets bigger.
- Later, we review calculus, in which we find the combined effect of \( x \), the slope of \( y \) as a function of \( x \) is

\[ \frac{dy}{dx} = f'(x) = \beta_1 + 2\beta_2 x \]  \hspace{1cm} (4)
The quadratic equation has one “turning point”.

A cubic equation can have (at most) two turning points.

The major problem with this kind of formula is that the tails—far left and far right—are very sensitive to changes in the coefficients for the squared and cubed terms.
Graph your own cubic

In R, run

```r
curve(20 + 0.08 * x + 0.015 * x^2 + 0.0002 * x^3, from = -100, to = 80)
```

Adjust the coefficients, you will notice the relationship is very sensitive.

- Wonder where the “peaks” and “valleys” are? It turns out (proof omitted) that they are the roots of this quadratic equation

\[0.08 + 0.30x + 0.0006x^2\]
The Exponent Can Be a Fraction

- \( x^{1/2} = \sqrt{x} \). Square root (\( \text{sqrt} \)) of \( x \).
  - Square root is a number that can be multiplied with itself to recover \( x \).
  - Square root of a negative number is undefined in elementary math (complex numbers allow it).
- \( x^{1/3} = \sqrt[3]{x} \). Cube root of \( x \).

How to they get that? You’d better go read a math book. Honestly, I just remember it.

Almost never am I forced to think about a value like

\[ x^{\frac{a}{b}} \]

but, if I do, I’ll go read a math book.

I should have mentioned, \( x^0 = 1 \). No matter what \( x \) is, \( x \) to the power of 0 is 1.
The Exponent Can be Negative

- The reciprocal of $x$
  \[ x^{-1} = \frac{1}{x} \quad (5) \]
- A negative exponent simply means the value inside the brackets is a divisor
- This fact makes some calculations in probability and Calculus much easier
- The most frequently occurring usage of this notation in statistics, by far:
  \[ \frac{1}{(1 + e^{-x})} = (1 + e^{-x})^{-1} \quad (6) \]
- Here, “$e$” is famous, Euler’s constant (don’t worry about that now)
- Publishers like the superscript $-1$ because it fits an equation into the one line of text
Compare sqrt and reciprocal

- $x^{1/2} = \sqrt{x}$
- $x^{-1} = \frac{1}{x}, x > 0$

**Notes about the reciprocal**
- $1/x$ is undefined if $x$ is 0
- As $x$ grows larger, $1/x$ tends toward 0. 0 is the “asymptotic value” of $\frac{1}{x}$.

**Notes about sqrt:** if $x < 0$, we have to change to “complex” math, which I find frustrating.
Summary of Things Worth Remembering about Exponents

- $x^{-b} = \frac{1}{x^b}$
- $x^a x^b = x^{a+b}$
- Too obvious? $x^3 x^2 = x \cdot x \cdot x \cdot x \cdot x = x^5$
- $x^a / x^b = x^{a-b}$ Consequence of previous 2 rules

- The value $b = 1$ is an important dividing point
- $b = 1$ implies $x^1$, just $x$
- $0 < b < 1$ implies a fractional power (remember square root)
- $1 < b$ implies something like $x$ squared.
- $b < 0$ means $\frac{1}{x^b}$
I’d forgotten that fractional exponents are defined only on the positive real numbers.

```r
curve(x^1, from = -5, to = 5)
curve(x^1.5, from = -10, to = 10, col = "green", add = TRUE)
curve(x^1.9, from = -10, to = 10, col = "red", lwd = 2, add = TRUE)
curve(x^2, from = -10, to = 10, col = "black", lwd = 1, add = TRUE)
```
Outline

1. Linear relationships
2. Nonlinear Overview
3. Splines and Segments
4. Polynomials
5. Exponentials
6. Logarithms
Something, to the power of $x$

- Before, we considered $x$ to the power of something
- Now, we consider something raised to the power of $x$
- As seen on right, generally, the value is very small on left, very huge on right.
Laws of Exponents (example with base 2)

\begin{align*}
2^0 &= 1 \\
2^1 &= 2 \\
2^{-1} &= 1/2 \\
2^{-\text{anything}} &= 1/2^{\text{anything}} \\
2^{\text{anything}} &\text{ is always positive, whether} \\
\text{anything} &\not\leq 0 \\
\text{All of these properties hold for other bases besides 2, of course (nonzero real numbers, positive or negative).}
\end{align*}

\begin{align*}
2^{3+5} &= 2^32^5 \quad \text{(for any a, b, } 2^{a+b} = 2^a2^b) \\
2^{3/2} &= (2^{1/2})^3 \\
\left(\frac{2}{k}\right)^3 &= \left[\frac{2^3}{k^3}\right], \quad k \neq 0
\end{align*}
Euler’s Constant (pronounced “Oiler’s”)

- Euler’s constant is rather like $\pi$ in the calculation of the area of a circle. It “just is”! It makes the math work out right.
- The best explanation of Euler’s constant I know of is the following. Consider a sequence.

$$1 + \frac{1}{1} + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4} + (\text{continue forever})$$

Euler proved that infinite sum is finite! It converges to a value, approx 2.71, which we call Euler’s constant, $e$.
- It has a number of seemingly magic properties. It shows up in many contexts.
- Because publishers want to save space, authors are pressured to replace

$$e^x$$

in their manuscripts with

$$\exp(x)$$
Exponential Relationships are Explosive

Toward the end of the first year calculus course, they ask

how many rabbits will you have if you start today with 10 rabbits and wait t years, assuming that the number of rabbits increases at a rate of 20 percent per year. (That’s an unrealistically low estimate, incidentally)

A lady bunny can create 4-8 more bunnies every 3 months or so.

This is motivated by a true story. Thomas Austin released “about 20” bunnies in the wild of Australia in 1859. 10 years later, there were several million rabbits (Wikipedia claim: 2 million were harvested in 1869 and their numbers were barely dented). Speculation is that the population grew at 150% per year for more than 10 years!
Exponential Growth of Bunnies

- After a cloud of dust, you find out that the number of rabbits will be the initial population times a constant e that is raised to a power that depends on the growth rate

\[ 10e^{0.20t} \]

*Again, there’s Euler’s constant, e is approx 2.71…*

- So the answer has
  - 10: initial population
  - Growth through t years: \( e^{0.20t} \)
Plot .20 Population Growth Rate (120% pop expansion)

In R, try

```R
curve(10*exp(0.20*x), from = -2, to = 10, xlab = "x represents time", ylab = "bunnies!")
```
Plot .20 Population Growth Rate (120% pop expansion)

In R, try

```r
curve(10*exp(0.20*x), from = -2, to = 10, xlab = "x represents time", ylab = "bunnies!")
```
Throw in .30 growth rate (130% per year) for comparison.

\[
\text{curve}(10*\exp(0.20*x), \text{from} = -2, \text{to} = 10, \text{xlab = "x represents time"}, \text{ylab = "bunnies!"}, \text{ylim = c(0, 200)})
\]
\[
\text{curve}(10*\exp(0.30*x), \text{from} = -2, \text{to} = 10, \text{xlab = "x represents time"}, \text{ylab = "bunnies!"}, \text{add = TRUE, col = "red"})
\]
If Australian Climate Really Did allow 120% Growth
In observed data, we often find only positive numbers.

We often build a theory using an exponential to enclose everything, so that our model makes positive predictions.

\[ y = e^{\beta_0 + \beta_1 x} \]

As long as the theory has \( \beta_1 > 0 \), then it is easy to visualize that (see previous page).
What if the argument is negative?

\( \exp(-x) \) is a ski-slope! A mirror image

Run this

```r
curve(10*exp(-1.02*x), from = -5, to = 5, xlab = "x represents time", ylab = "bunnies!")
```
What if the argument is negative?

\( \exp(-x) \) is a ski-slope! A mirror image

- Run this

```r
curve(10*exp(-1.02*x), from = -5, to = 5, xlab = "x represents time", ylab = "bunnies!")
```

- See?
The Importance of Remaining Positive

- We have a lot of research problems in which ONLY positive values are observed.
- To build a predictive theory for that kind of data, BY FAR the most common approach is to suppose that
  \[ y = e^{\beta_0 + \beta_1 x} \]

- Linear Predictor: Term for the exponent, \( \beta_0 + \beta_1 x \).
- They often log both sides of the equation, so that the linear predictor is by itself
  \[ \log(y) = \beta_0 + \beta_1 x \]

- That thing on the left is commonly called the “link function” (see “Generalized Linear Models”)
- In order to explain that, we need to discuss logarithms, where are next! (note clever transitional ploy)
Outline

1. Linear relationships
2. Nonlinear Overview
3. Splines and Segments
4. Polynomials
5. Exponentials
6. Logarithms
Logs result from thinking backwards

- Recall that the square root requires us to think backwards? $\sqrt{81}$ asks us to think “what number would we need to multiply by itself to obtain 81?”

- The logarithm is a similar sort of backward thinking. The expression $\log_{10}100$ means “to what power must we raise 10 to obtain 100”. Obviously, the answer is 2. (We hope $10^2 = 100$).

- Think through this one. What is $\log_{10}1000$
Logs result from thinking backwards

- Recall that the square root requires us to think backwards? $\sqrt{81}$ asks us to think “what number would we need to multiply by itself to obtain 81?”

- The logarithm is a similar sort of backward thinking. The expression $\log_{10}100$ means “to what power must we raise 10 to obtain 100”. Obviously, the answer is 2. (We hope $10^2 = 100$).

- Think through this one. What is $\log_{10}1000$?

  $\log_{10}1000 = 3$
Various bases for logs

- The subscript on the log is called the *base*.
- The base can be any positive number. Base 10 is most understandable to us because we have learned math in a base 10 framework since childhood.
- In computer science, where values are often stored in little bits that are either On or Off, they often use base 2.
- What is:

  \[ \log_2 4 \]
Plot some Logs

Please try this in R

```r
curve(log(x, base = 10), from = 0.0001, to = 4)
curve(log(x, base = 2), from = 0.0001, to = 4, add = TRUE)
curve(log(x), from = 0.001, to = 4, main = "If you omit the base, you get base = Euler's constant")
```
Compare some Logs with various bases

Did you notice?

- \( \log_b(1) = 0 \). The log of 1 is 0, no matter what the base.
- \( \log_b(0) \) is undefined. As \( x \) goes to 0, \( \log_b(0) \) tends to \(-\infty\).
- \( \log \) is not defined for negative numbers either.
Euler’s constant as the base

- In mathematical applications, Euler’s constant is the most frequently used base.
- One of the reasons for that is this peculiar fact. The slope of the function $\log_e(x)$ is $1/x$.
- If you consider any other base, then we have to remember to include a correction factor in the slope (and that’s annoying).
- Because the log with the base $e$ has that convenient property, it is called the “natural log” and often it is called $\ln(x)$.
- In R, the function log() defaults to the base $e$. Other programs differ, perhaps calling it “loge” or such.
**exp and ln are inverses of each other**

- *ln* reverses the effect of *exp*, meaning \( \ln(\exp(x)) = x \)
  
  Suppose 
  \[
  z = \exp(x) = e^x
  \]
  
  Then \( \ln(z) = x \).

  Why? just write it in:
  \[
  \log_e(e^x)
  \]

  We are asking “to what power must we raise \( e \) in order to obtain \( e^x \)”.
  Obviously, the answer is \( x \).

exp reverses ln

- For some reason, this one is more difficult for me to visualize.
- People always write \( \exp(\ln(x)) = x \), I always believed that was true.
- I want to convince you (and me!)

**Theorem**

If \( z = \ln(x) = \log_e(x) \), then \( \exp(z) = x \)

- Muddle through a proof of that
  - If \( z = \ln(x) \), that means \( e^z = \exp(z) = x \).
  - Then \( e^z = e^{\log_e(x)} \) (and for reasons I cannot put into words) = x
Various Reasons to use Logs

1. Theoretically meaningful
2. Simplify formulae
3. Keep numbers within manageable limits
1: Theoretically Meaningful

- Sometimes we suspect that $y$ is increasing more-and-more-slowly as $x$ increases

  \[ y = \beta_0 + \beta_1 \log(x_i) \]

- As wealth increases, happiness increases more and more slowly

- As exercise increases, a person’s fitness increases more and more slowly
Logs have some weird properties that make them very useful.

1. “The log of a product is the sum of the logs”
   \[ \log(x_1 \cdot x_2) = \log(x_1) + \log(x_2) \]
   
   \[ \log(x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_m) = \log(x_1) + \log(x_2) + \log(x_3) + \ldots + \log(x_m) \]

Why is that useful?

- In statistics and probability, we often end up with a lot of terms multiplied together. A big tangled product is almost always mathematically unworkable. So we log it to convert it to the sum of separate pieces.
- Keep your eye out for “log likelihood” as you move forward.
2: Simply formulae

2 “The log of $x^{\text{anything}}$ is $\text{anything} \times \log(x)$”

Why is that useful?

- We often find a statistical quantity that looks like this
  \[ p_1^\alpha \cdot p_2^\beta \cdot p_3^\gamma \]

- Apply the first rule
  \[ \log(p_1^\alpha \cdot p_2^\beta \cdot p_3^\gamma) = \log(p_1^\alpha) + \log(p_2^\beta) + \log(p_3^\gamma) \]

- Apply the second rule
  \[ = \alpha \log(p_1) + \beta \log(p_2) + \gamma \log(p_3) \]
2: Simplify formulae

3 “The log of a fraction is the difference of the logs”

\[ \log(x_1/x_2) = \log(x_1) - \log(x_2) \]

This rule can be deduced from the previous rules. To see why, write it as

\[ \log(x_1 \cdot x_2^{-1}) \]

Apply rule 1. Then rule 2.

\[ \log(x_1) + \log(x_2^{-1}) = \log(x_1) - \log(x_2) \]
3: Keep numbers in manageable limits

- Sometimes we collect numbers on huge scales
- Often they have plenty of values clustered in a low range, but then there are less frequent, extreme observations
- It is common practice to take the log of those variables and then put them to use "as if" the logged variable were the truly interesting one.
3: Animal brains!

- To appreciate the importance of this point, we would have to step ahead one or two semesters into applied statistics.
- But I can’t resist the temptation to show you an example.
- In the MASS package for R, there is a dataset called Animals. It has 2 variables, an animal’s body weight and the weight of its brain.
3: Animal brains!

Histogram: Brain Weight of Animals

Density

Animals$brain
3: Animal brains!

Scatterplot: Brain and Body Weight

- The scatterplot shows the relationship between brain weight and body weight across different animals.
- The x-axis represents body weight, while the y-axis represents brain weight.
- There are a few data points indicating the brain-to-body weight ratio for different species.
3: Animal brains!

Scatterplot: Brain and Body Weight

Body Weight (natural log)

Brain Weight (natural log)
3: Animal brains!

Scatterplot: Brain and Body Weight

Brain Weight (natural log)

Body Weight (natural log)

Diplodocus

Triceratops

Brachiosaurus
I’ve had logs on the brain lately

- We often have data in proportions or percentages.
- It seems obvious, but this simple problem was overlooked in statistics for about 100 years: percentages do not behave like other numbers.
- The percentage of the population that is literate increases, but as the percentage approaches an “upper limit”, it grows more and more slowly.
- The percentage of children with measles may decline from year to year, but as it approaches 0, it declines more and more slowly.
A common translation for proportions

- This is very widely used when data is thought of as a proportion

$$\ln \left[ \frac{P}{1-P} \right]$$

- That “logit transformation” converts input on the 0-1 scale into a variable that varies from minus to plus infinity.
A common translation for proportions ...

- Ways to think about the logit transformation.
  - \( P/(1 - P) \) is the “odds ratio”. If \( P = p_1 \) is the proportion in category 1, and the other outcome is \( p_2 = 1 - p_1 \), then \( p_1/p_2 \) can get as small as 0 and as big as infinity.
A common translation for proportions ...

- Log that to turn it into a number that ranges from $\infty$ to $\infty$
- Can I show you something interesting? Suppose

$$\ln \left[ \frac{P}{1-P} \right] = \beta_0 + \beta_1 x$$

then

$$P = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

which is the same as

$$P = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$$

and because publishers hate to waste lines, they’ll often write

$$P = \frac{1}{1 + \exp(-(\beta_0 + \beta_1 x))} = \left(1 + \exp(-(\beta_0 + \beta_1 x))\right)^{-1}$$
You want me to prove that? You try first

Start Here

$$\ln \left[ \frac{P}{1 - P} \right] = \beta_0 + \beta_1 x$$

$$\frac{P}{1 - P} = \exp(\beta_0 + \beta_1 x)$$

And rearrange several times
You want me to prove that? You try first ...

\[ P = \exp(\beta_0 + \beta_1 x) - P \exp(\beta_0 + \beta_1 x) \]

\[ P + P \exp(\beta_0 + \beta_1 x) = \exp(\beta_0 + \beta_1 x) \]

\[ P(1 + \exp(\beta_0 + \beta_1 x)) = \exp(\beta_0 + \beta_1 x) \]

\[ P = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \]

\[ P = \frac{\exp(\beta_0 + \beta_1 x) / \exp(\beta_0 + \beta_1 x)}{1 / \exp(\beta_0 + \beta_1 x) + \exp(\beta_0 + \beta_1 x) / \exp(\beta_0 + \beta_1 x)} \]

\[ P = \frac{1}{1 / \exp(\beta_0 + \beta_1 x) + 1} \]

\[ = \frac{1}{\exp(-(\beta_0 + \beta_1 x)) + 1} \]

\[ = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \]
Extend that to compositions of more than one type

- We often have data that is “compositional”.
- John Aitchison (famous 1986 book *Analysis of Compositional Data*).
- If the proportions of a whole are \((p_1, p_2, p_3, p_4)\) we have various transformations we can employ.
- Suppose we make category 4 the baseline category, then the proportion provides only enough information to calculate 3 proportional log ratios:
  \[
  \ln \left( \frac{p_1}{p_4} \right), \quad \ln \left( \frac{p_2}{p_4} \right), \quad \ln \left( \frac{p_3}{p_4} \right)
  \]
- There are several theorems in Aitchison’s book which claim that any valid analytical conclusion based on this transformation must also hold true if we put the other proportions in as the base.
Allegedly, Analysis Follows

- There strong claim in the compositional data field is that, after making that “logistic normal” transformation, then we proceed to data analysis as if those numbers were “multivariate, related observations” on separate variables.

- More likely, you will run into this at the end of the first class in regression analysis, or possibly in a follow-up course that focuses on categorical data analysis.