Day 4: Linear Algebra and Calculus

Even more unsolicited advice

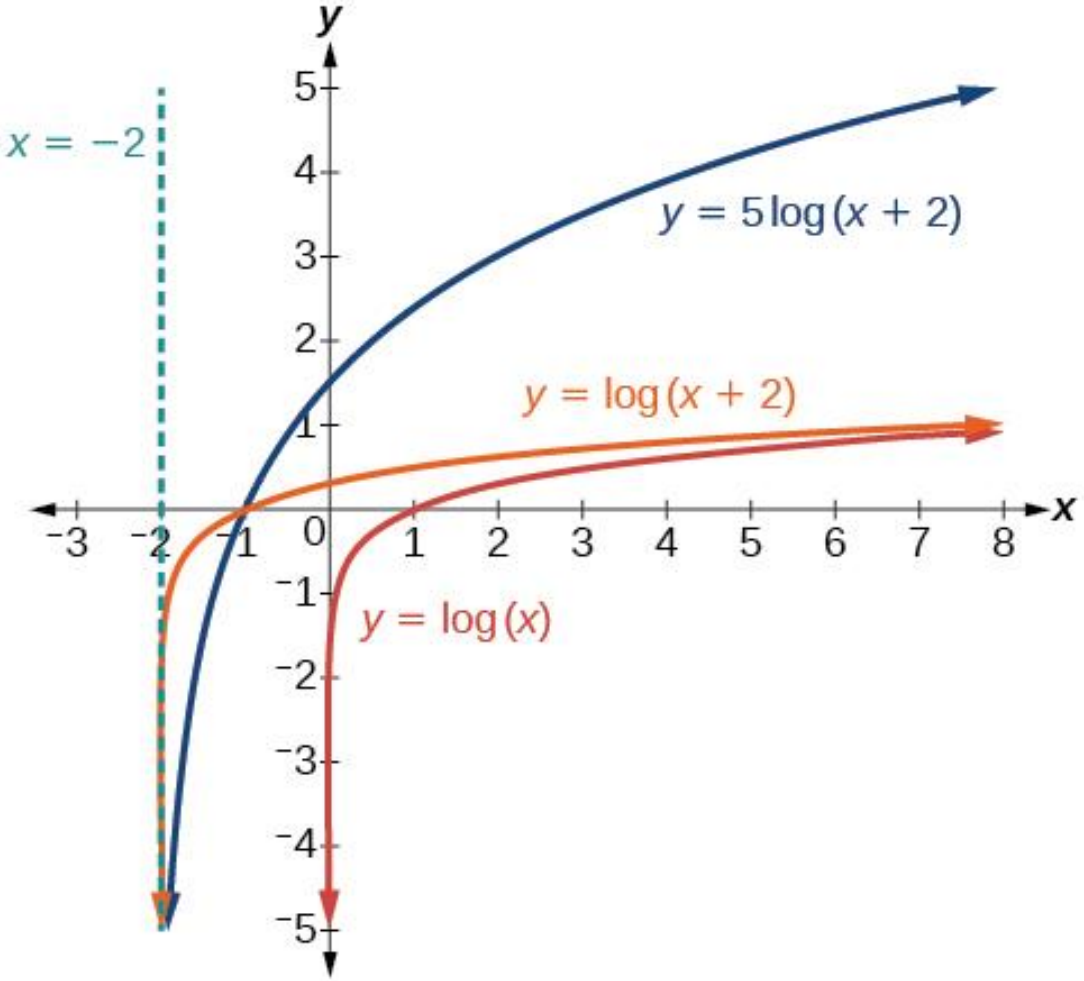
- Monitors:
  - At least 1920x1080p resolution
  - 16:9 screen ratio
  - At least 60hz refresh rate
  - Max 10ms response time
  - IPS is a good all around panel.
  - 24” minimum, 27” is the sweet spot for most people
  - Check what your computer can support, connections and number.
- Yes your computer is powerful enough (probably)
- Statistics is hard

Review: Thinking geometrically

- If I subtract X from Z, what does the resulting value tell me about the relationship between the two numbers?

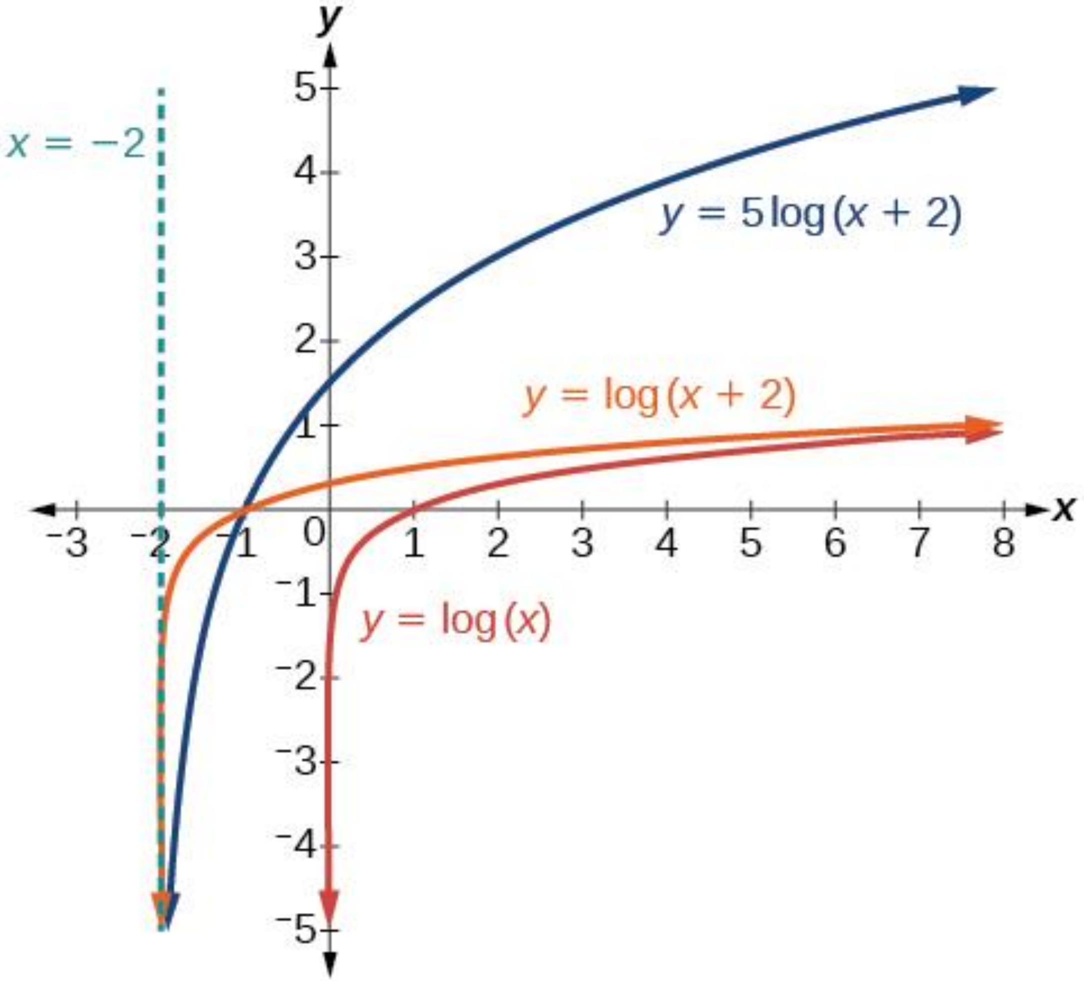
Review: Logarithms

- Why does log(x) asymptotically approach zero?



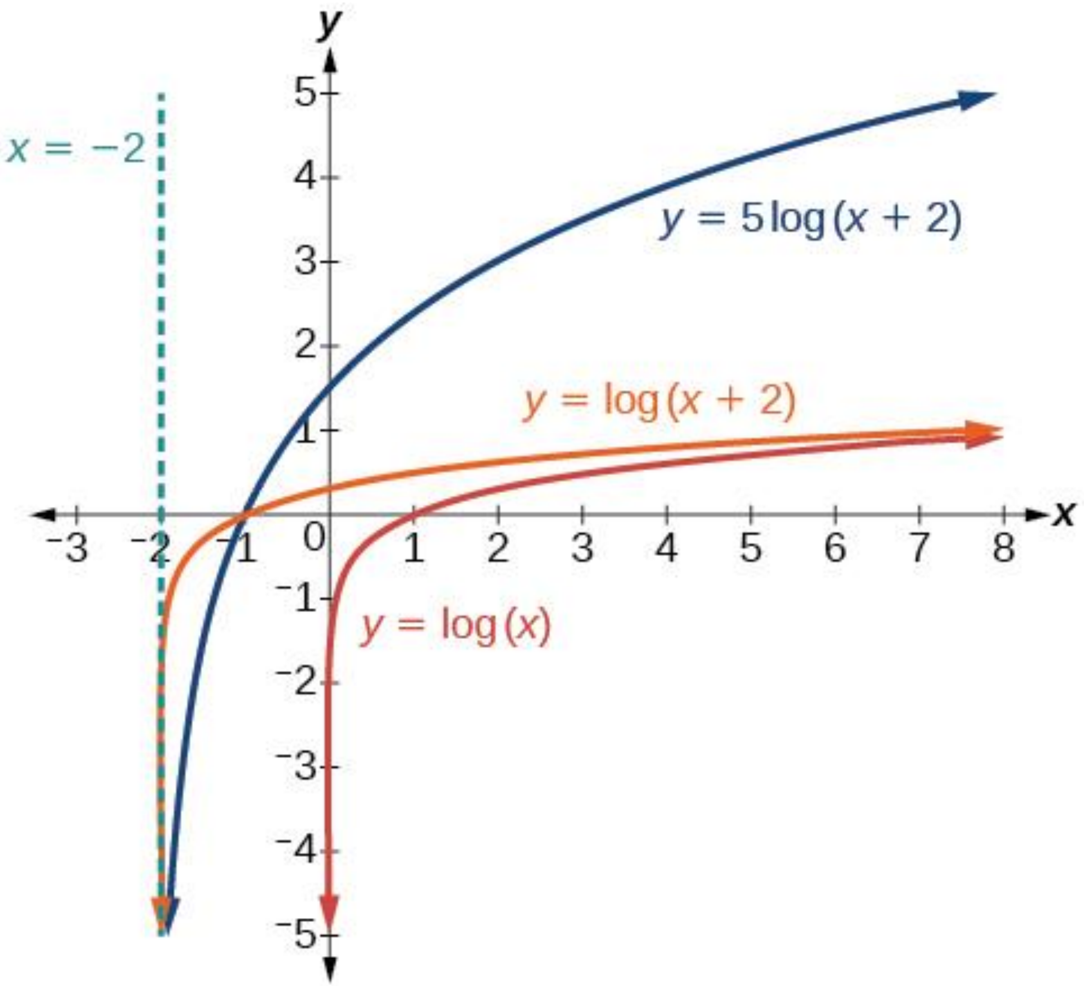
Review: Logarithms

- Why does log(x) asymptotically approach zero?
- Why is log(x+2) shifted to the left?



Review: Logarithms

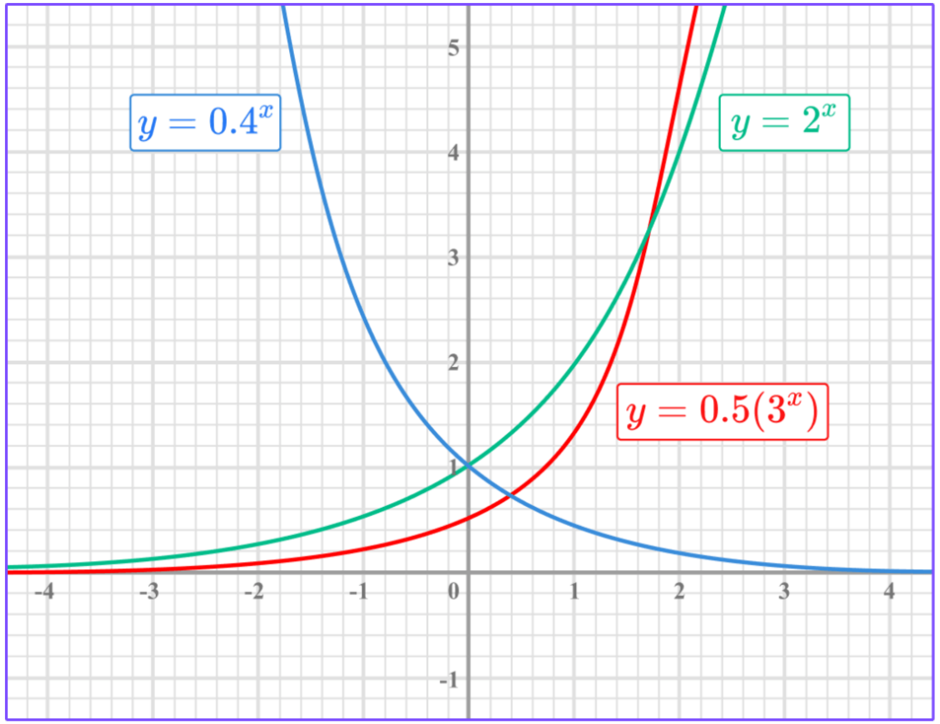
- Why does log(x) asymptotically approach zero?
- Why is log(x+2) shifted to the left?
- Why is 5log(x+2) less flat than the other lines?



Review: Exponential Functions

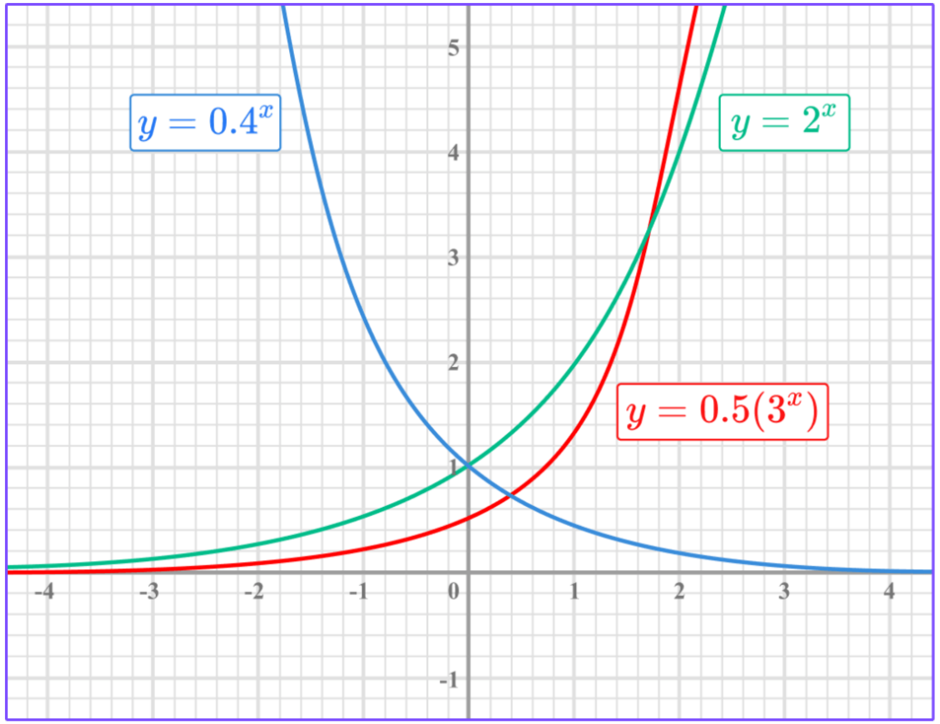
- The opposite of a logarithm
- When the variable is in an exponent
- $f(x) = a^x$

Why are each of these lines trending the way they are?



Review: Exponential Functions

- The opposite of a logarithm
- When the variable is in an exponent
- $f(x) = a^x$
- Remember division is multiplication by reciprocals



Functions

I have data on 1,000 people.

- *y*: if they voted in the most recent election
- *x*: their political ideology which is a continuous value between -2 and +2
- *c:* the travel time to their nearest polling station.

We've learned at least two different ways that we can express such a model.
What are they and what are the different assumptions that they imply?

One solution

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

What assumptions is this equation making about the relationship between the variables and Y?

(hint: think about what it would look like if you plotted one of the variables against Y)

How can we change the equation if we want to change our assumptions?

Another solution

$$logistic(Y_i) = \frac{1}{1 + e^{-(Y_i)}}$$

$$logistic(Y_i) =$$

Another solution

$$logistic(Y_i) = \frac{1}{1 + e^{-(Y_i)}}$$

$$logistic(Y_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 X_2 + u_i)}}$$

Goals

- Develop a conceptual understanding of derivatives and integrals
- Understand basic linear algebra operations and how they relate to computing

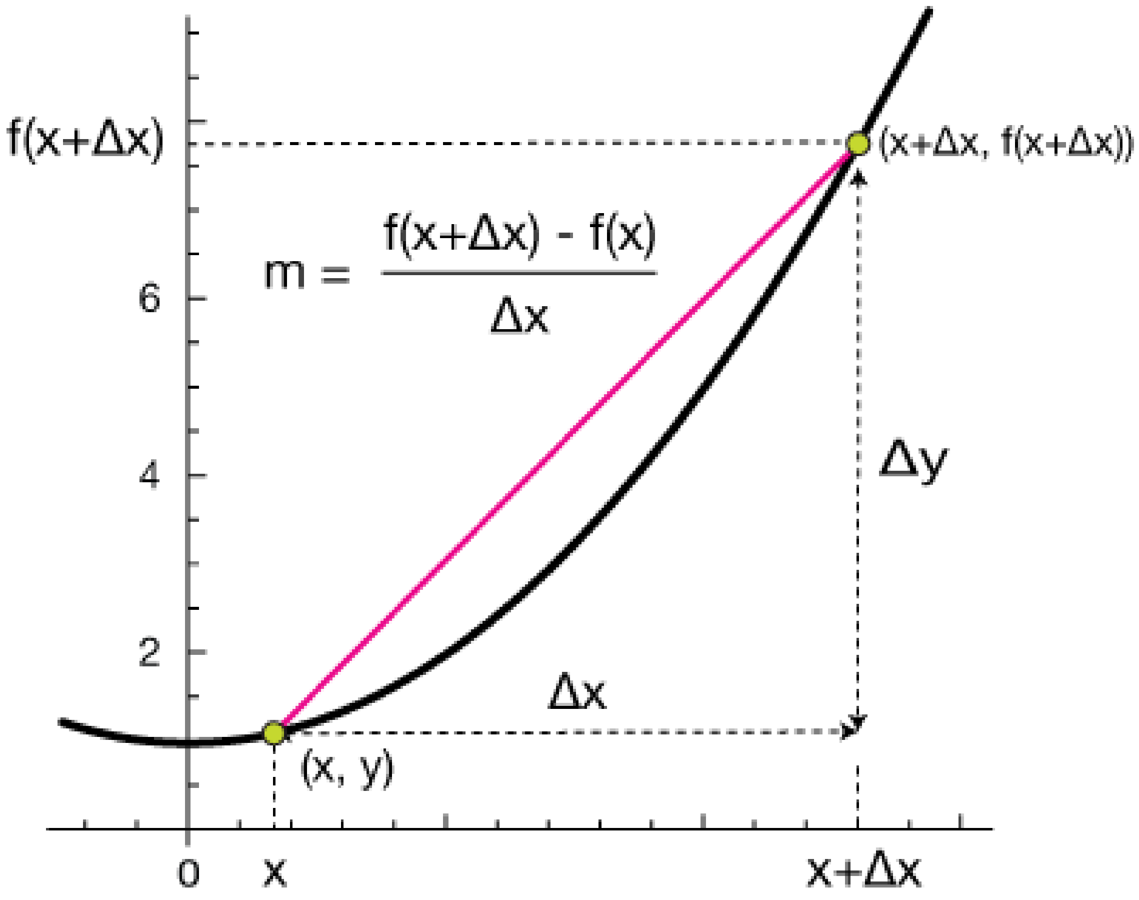
Derivatives

Limits

- Coefficient = Slope
- Slope = rate of change

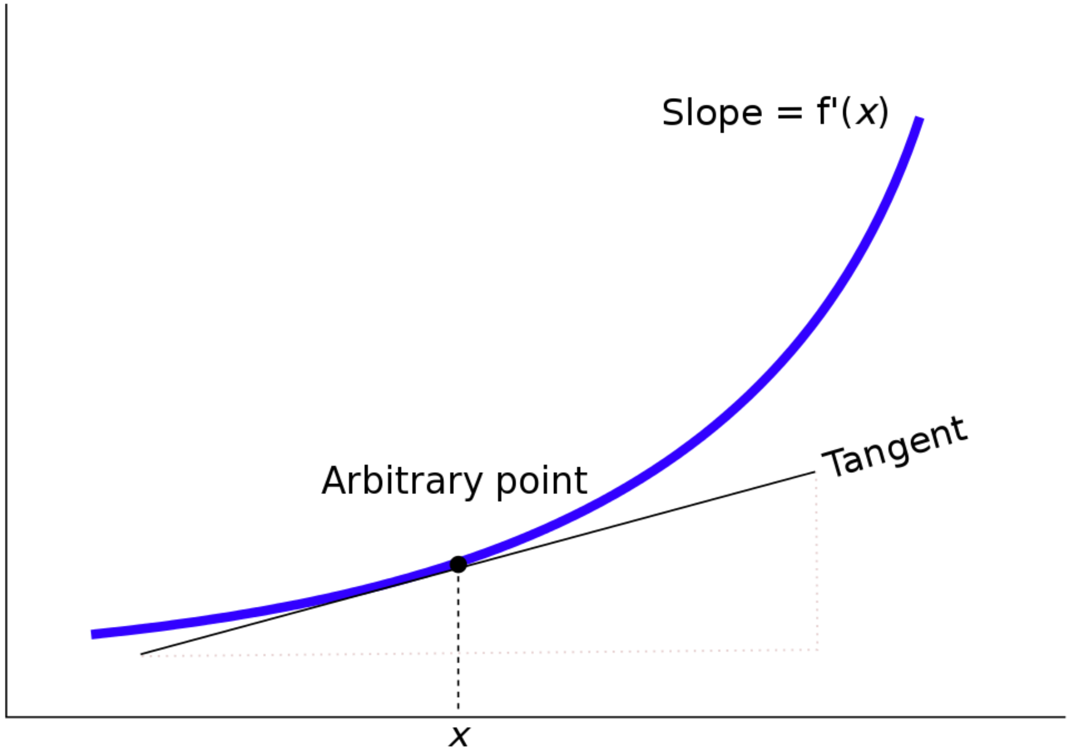
Limits

- Coefficient = Slope
- Slope = rate of change
- How do we determine what the rate of change is at a point on this curve?



Derivatives

- The slope of the line tangent to *f()* at *x*

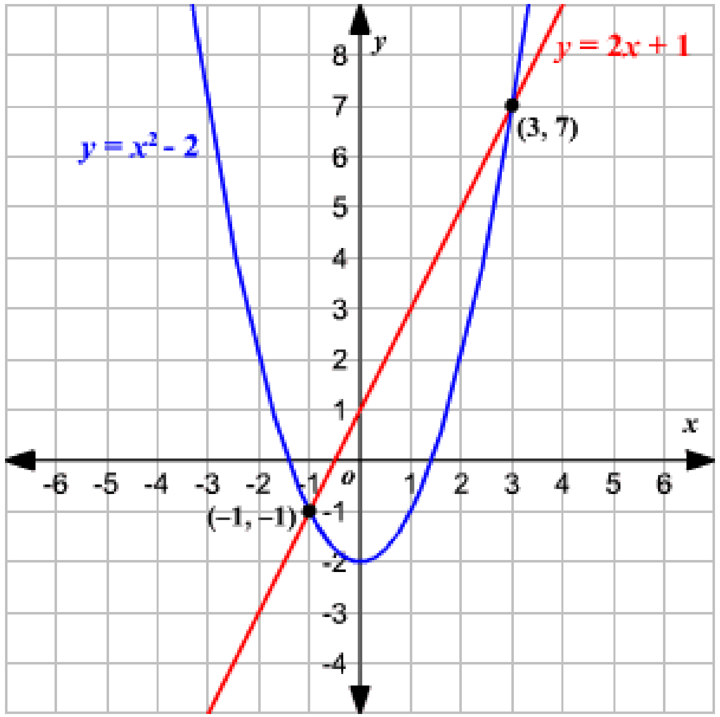


Why would we want to take a derivative in social science?

Why would we want to take a derivative in social science?

Two common examples:

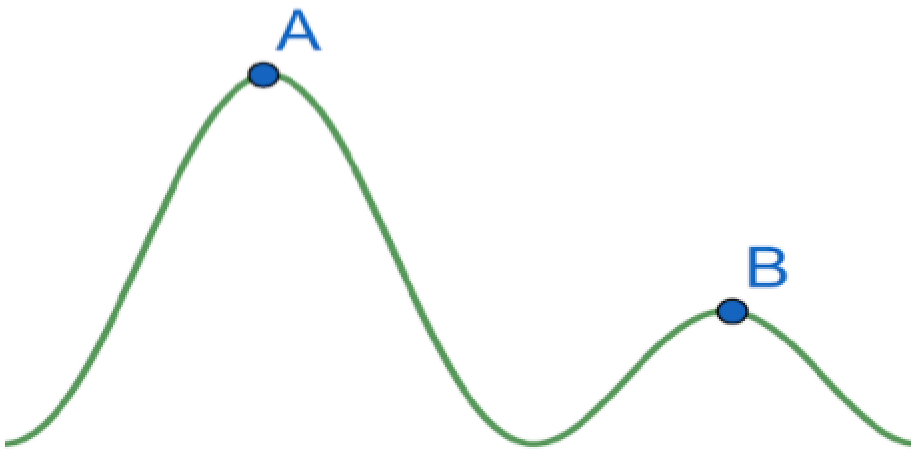
1. Estimating effect sizes on for non-linear equations



Why would we want to take a derivative in social science?

Two common examples:

1. Estimating effect sizes on for non-linear equations
2. Finding local minima and maxima of a distribution



Calculating derivatives (in brief)

- The power rule will get you 90% of the way there
- There are more general rules for all types of functions but we won't be going over those for now.
- Derivative of a constant = 0

Power Rule

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

$$\frac{d}{dx}(x) = 1$$

Partial derivatives

- If a function has more than one variable, you can take a derivative with respect to a single variable
- Useful when working with vectors

$$z = f(x, y) = x^2 + xy + y^2 .$$

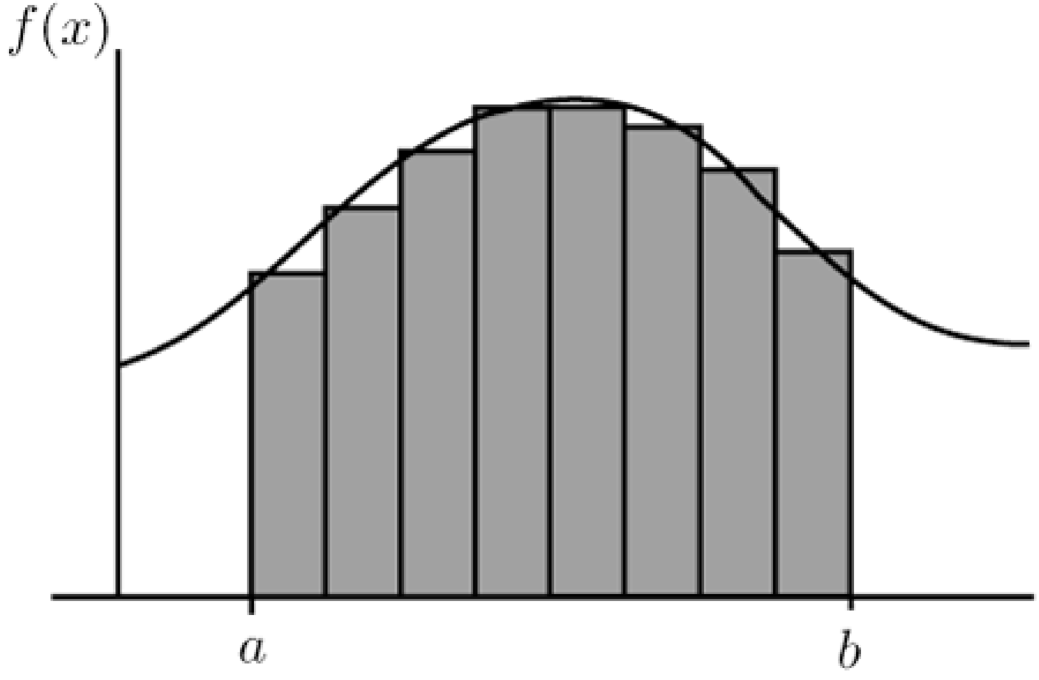
$$\frac{\partial z}{\partial x} = 2x + y.$$

Notation

	Lagrange	Leibniz
Function	$f(x)$	f
Derivative	$f'(x)$	$\frac{df}{dx}$
Second Derivative	$f''(x)$	$\frac{d^2 f}{dx^2}$
Higher Derivative	$f^{(n)}(x)$	$\frac{d^n f}{dx^n}$

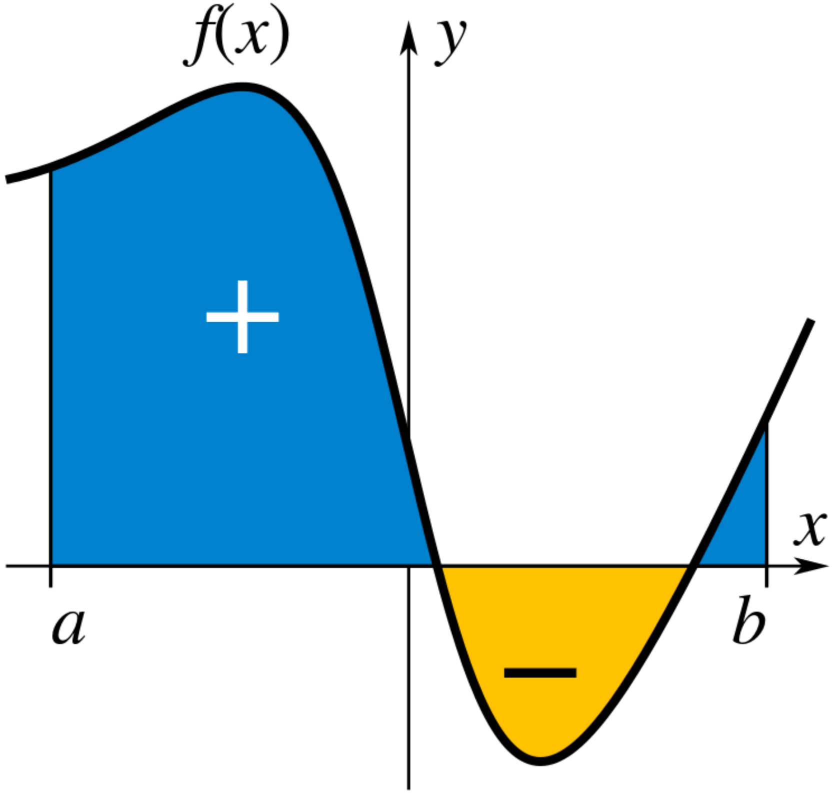
Integrals

Integrals in concept



Integrals

- Integrals are a continuous sum used to calculate the area under a curve
- Definite integrals calculate the area between two points.
- Indefinite integrals are functions whose derivatives are $f(x)$
- Differentiation and integration are inverse functions



Notation

Definite

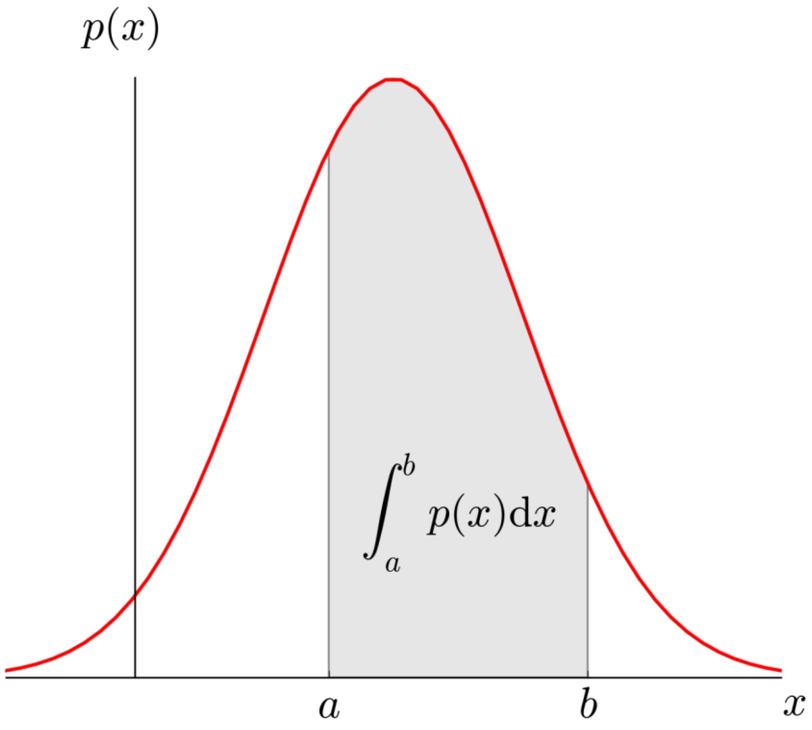
$$\int_a^b f(x)dx$$

Indefinite

$$\int f(x)dx$$

Applications of integrals

- When you think integration, you should think area and volume
- Integration is most commonly used in probability
- A probability is an area of a probability density function



Linear Algebra: Vectors

Linear algebra

- Solving systems of linear equations
- Arithmetic with vectors and matrices
- Very widely used in data science
- Particularly important in machine learning
- You won't often be doing linear algebra, but the models and algorithms you work with will

A quick refresher…

$$\begin{bmatrix} -85 & -55 & -37 & -35 & 97 \\ 50 & 79 & 56 & 49 & 63 \\ 57 & -59 & 45 & -8 & -93 \end{bmatrix}$$

3 × 5 matrix

$$\begin{pmatrix} 92t & 43 & -62 & 77 \\ 0 & 54 & -5t & 99 \\ 0 & 0 & -12 & -18t \end{pmatrix}$$

3 × 4 matrix

$$\begin{bmatrix} -85 \\ 50 \\ 57 \end{bmatrix}$$

3 × 1 matrix (a column vector)

$$\begin{pmatrix} 92 & 43 & -62 & 77 \end{pmatrix}$$

1 × 4 matrix (a row vector)

Basic operations

$$u = [3,3,3,3] \text{ and } v = [1,2,3,4]$$

$$u + v = [4, 5, 6, 7]$$

$$u \ - v = [2,1,0,-1]$$

Scalar multiplication and division

$$u = [3,3,3,3] \text{ and } v = [1,2,3,4]$$

$$3 \times u = [9, 9, 9, 9]$$

$$v \div 3 = \left[\frac{1}{3}, \frac{2}{3}, 1, \frac{4}{3}\right]$$

Dot product (inner product or scalar product)

- A product of two vectors that results in a scalar
- Most common application is to measure similarity. We won't get in to the how.

$$u = [3,3,3] \text{ and } v = [1,2,3]$$

$$\mathrm{u} \cdot v = [3 \times 1 + 3 \times 2 + 3 \times 3] = [3 + 6 + 9] = 18$$

Linear Algebra: Matices

- An $n \times m$ collection of vectors

Addition and Subtraction

Add and subtract matrices

$$\begin{bmatrix} 8 & 5 \\ 2 & 3 \end{bmatrix} + \begin{bmatrix} 9 & 5 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 17 & 10 \\ 3 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 8 & 5 \\ 2 & 3 \end{bmatrix} - \begin{bmatrix} 9 & 5 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}$$

- Two matrices must have the same dimensions
- Addition is Commutative

Multiplication by a scalar

$$2 \cdot \begin{bmatrix} 10 & 6 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 \cdot 10 & 2 \cdot 6 \\ 2 \cdot 4 & 2 \cdot 3 \end{bmatrix}$$

Multiplication by a vector

Dot product

$$\begin{pmatrix}1&1&2\\2&1&3\\1&4&2\end{pmatrix}\begin{pmatrix}3\\1\\2\end{pmatrix} = \begin{pmatrix}1\cdot3 + 1\cdot1 + 1\cdot2\\ \\ \end{pmatrix}$$

First row,

$$\begin{pmatrix}1&1&2\\2&1&3\\1&4&2\end{pmatrix}\begin{pmatrix}3\\1\\2\end{pmatrix} = \begin{pmatrix}1\cdot3 + 1\cdot1 + 1\cdot2\\ 2\cdot3 + 1\cdot1 + 3\cdot2\\ \end{pmatrix}$$

next row,

$$\begin{pmatrix}1&1&2\\2&1&3\\1&4&2\end{pmatrix}\begin{pmatrix}3\\1\\2\end{pmatrix} = \begin{pmatrix}1\cdot3 + 1\cdot1 + 1\cdot2\\ 2\cdot3 + 1\cdot1 + 3\cdot2\\ 1\cdot3 + 4\cdot1 + 2\cdot2\end{pmatrix} = \begin{pmatrix}6\\13\\11\end{pmatrix}$$

last row,
then do the addition.

Multiplication by a matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \text{ x } \begin{bmatrix} 10 & 11 \\ 20 & 21 \\ 30 & 31 \end{bmatrix}$$

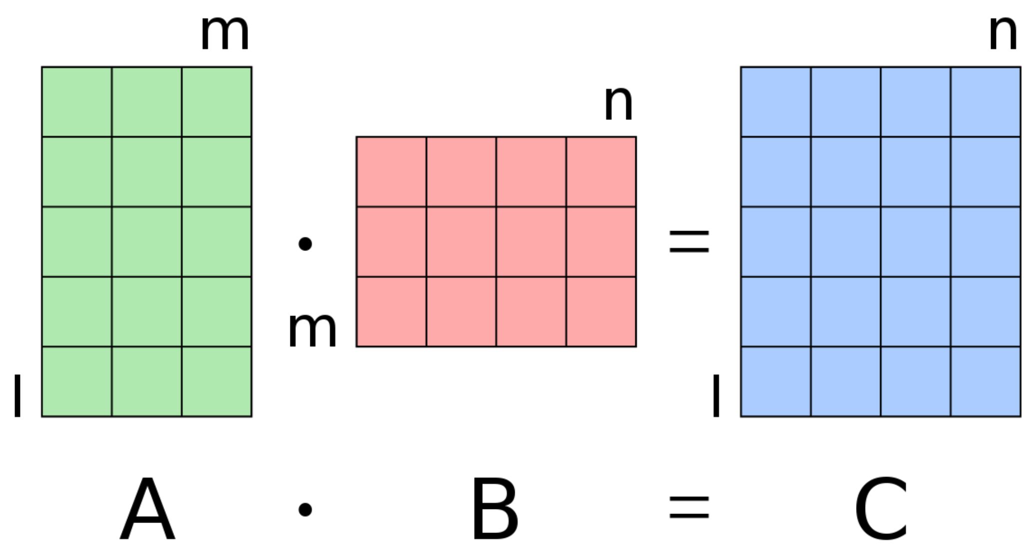
Dot product

$$= \begin{bmatrix} 1\text{x}10 + 2\text{x}20 + 3\text{x}30 & 1\text{x}11 + 2\text{x}21 + 3\text{x}31 \\ 4\text{x}10 + 5\text{x}20 + 6\text{x}30 & 4\text{x}11 + 5\text{x}21 + 6\text{x}31 \end{bmatrix}$$

$$= \begin{bmatrix} 10+40+90 & 11+42+93 \\ 40+100+180 & 44+105+186 \end{bmatrix} = \begin{bmatrix} 140 & 146 \\ 320 & 335 \end{bmatrix}$$

Matrix multiplication rules

- Rows × Columns
- The first matrix must have the sane number of columns as the second matrix has rows.
- Not commutative



Transposing matrices

- Just think of it as swapping rows and columns
- When you multiply a matrix A by its transpose A^T the resulting matrix is always a square matric

$$\begin{bmatrix} 1 & 2 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

Identity matrix

- Denoted as $I_{n \times n}$
- A square matrix in which the diagonal consists of ones and the rest of the values are zero.
- Any matrix multiplied by an identity equals itself

$$I_1 = 1$$

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Matrix Inverse

- The inverse of matrix A is A^{-1} when:

$$AA^{-1} = A^{-1}A = I$$

- Only square matrices have an inverse

Matrix expression of a model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + \ u_i$$

$$y = \boldsymbol{X\beta} + \boldsymbol{e}$$

Solving linear regression

We want to solve for $\boldsymbol{\beta}$ in the following equation:

$$y = \boldsymbol{X\beta}$$

But we can't divide by matrices, we multiply by their inverse instead. We can't invert a matrix that isn't square, so we start by multiply both sides of the equation by $\boldsymbol{X}^T$.

$$\boldsymbol{X}^T y = \boldsymbol{X}^T \boldsymbol{X\beta}$$

Now that we have a square matrix on the right side, we can multiply by the inverse of $\boldsymbol{X}^T\boldsymbol{X}$ to isolate $\boldsymbol{\beta}$:

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T y = \boldsymbol{\beta}$$

This is called the "normal" equation. You don't really need to know this.