Intro to intro to statistics

Frank Edwards

9/6/21

Welcome to Graduate statistics training!

Stats camp goals

1. Get your software set up
2. Remind you what these things mean:

$$log(x)$$

$$y = 2 + 3x$$

$$\frac{d}{dx}x^3 = 3x^2$$

$$\sum_{x=1}^{10} x$$

Today: Let's install software

Step 1

Install R. `https://cran.r-project.org/`

Step 2

Install RStudio. `https://www.rstudio.com/`

Today: Let's install software

Step 1

Install R. `https://cran.r-project.org/`

Step 2

Install RStudio. `https://www.rstudio.com/`

- Does anyone have problems with regular access to a computer that can run R?

Problems?

Getting started in RStudio

- The script and the console

Getting started in RStudio

- The script and the console
- Make it pretty! (You will be spending lots of time here...)
  - Themes
  - Pane layouts

Getting started in RStudio: packages

- Install packages

```
install.packages("tidyverse")
```

Getting started in RStudio: packages

- Load packages

```r
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ ti
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Problems?

Using R as a (very fancy) calculator

```r
2 + 2
```

Using R as a (very fancy) calculator

```
2 * 2
```

Using R as a (very fancy) calculator

```
2 / 2
```

Using R as a (very fancy) calculator

```
2 ^ 3
```

Using R as a (very fancy) calculator

```
sqrt(4)
```

Using R as a (very fancy) calculator

```r
# I am a comment! I help you read code!
# the <- operator makes assignments
# Make a new variable called 'x', set it equal to 2
x <- 2
x^2
```

Compute the following in R (with $x = 2$)

- x^3

Using R as a (very fancy) calculator

Compute the following in R (with $x = 2$)

- x^3
- $2x$

Compute the following in R (with $x = 2$)

- x^3
- $2x$
- $\frac{x}{2}$

Using R as a (very fancy) calculator

Compute the following in R (with $x = 2$)

- x^3
- $2x$
- $\frac{x}{2}$
- $(2 + x)^2$

Compute the following in R (with $x = 2$)

- x^3
- $2x$
- $\frac{x}{2}$
- $(2 + x)^2$

Before we proceed

- Are you running RStudio?

Before we proceed

- Are you running RStudio?
- Have you installed tidyverse?

Before we proceed

- Are you running RStudio?
- Have you installed tidyverse?
- When you type 2==3 in the console and hit enter, what does it say?

Before we proceed

- Are you running RStudio?
- Have you installed tidyverse?
- When you type 2==3 in the console and hit enter, what does it say?

Exciting! You are programming in R!

Break

Welcome back: math time!

Agenda today

1. Math notation and key concepts
2. Exponents, logarithms
3. Lines and graphs

Real numbers and integers

Real numbers (doubles)

- Any continuous number
- E.g. 4, 4.189, 2/3, π

Integers

- Any whole number
- 10, -10, 24, 87

Variables

- May take on any value
- Represented by letters, such as x, y, z
- Can be used in any mathematical operation

Functions

- Maps each element of set x to an element of set y
- Often denoted by f, g, h
- $f(x) = 2x + 3$

Parameters (statistics)

- Variable that represent a feature of a population
- Represented by Greek letters, such as μ, σ, ε

Summation

Represented as $\sum$, with integer begin and end points

$$\sum_{x=1}^{3} x$$

Summation

Represented as $\sum$, with integer begin and end points

$$\sum_{x=1}^{3} x$$

$$\sum_{x=1}^{3} x = 1 + 2 + 3 = 6$$

Summation in R

In R, we can calculate a sum using the sum() function

```
# make an integer vector from 1 to 3
x<-1:3
# x<-c(1,2,3) is equivalent
sum(x)
```

```
## [1] 6
```

Summation review

Compute the following by hand, and then in R

$$\sum_{x=3}^{8}(x+1)$$

$$\sum_{x=1}^{4} 2x$$

Compute the following in R

$$\sum_{x=3}^{8}(x+1)$$

$$\sum_{x=1}^{4}2x$$

Exercises (solutions)

Compute the following in R

$$\sum_{x=3}^{8}(x+1)$$

```r
x<-3:8 # or c(3, 4, 5, 6, 7, 8)
sum(x+1)
```

$$\sum_{x=1}^{4}2x$$

```r
x<-c(1, 2, 3, 4) # or 1:4
sum(2*x)
```

Products

Represented as $\prod$, with integer begin and end points

$$\prod_{x=1}^{4} x$$

Products

Represented as $\prod$, with integer begin and end points

$$\prod_{x=1}^{4} x$$

$$\prod_{x=1}^{4} x = 1 \times 2 \times 3 \times 4$$

Products

Represented as $\prod$, with integer begin and end points

$$\prod_{x=1}^{4} x$$

$$\prod_{x=1}^{4} x = 1 \times 2 \times 3 \times 4$$

In R:

```
x<-1:4
prod(x)
```

```
## [1] 24
```

Exponents and logarithms

Exponents

a^n Is equal to a multiplied by itself n times

Exponents

a^n Is equal to a multiplied by itself n times

- $2^3 = 2 \times 2 \times 2 = 8$

Exponents

a^n Is equal to a multiplied by itself n times

- $2^3 = 2 \times 2 \times 2 = 8$
- $5^4 = 5 \times 5 \times 5 \times 5 = 625$

In R, we can calculate exponents using the ^

```r
2^3
```

```
## [1] 8
```

```r
5^4
```

```
## [1] 625
```

Some rules of exponents

$$x^1 = x$$

Some rules of exponents

$$x^1 = x$$

$$x^0 = 1$$

Some rules of exponents

$$x^1 = x$$

$$x^0 = 1$$

$$x^k + x^l = x^{k+l}$$

Some rules of exponents

$$x^1 = x$$

$$x^0 = 1$$

$$x^k + x^l = x^{k+l}$$

$$(x^k)^l = x^{kl}$$

Some rules of exponents

$$x^1 = x$$

$$x^0 = 1$$

$$x^k + x^l = x^{k+l}$$

$$(x^k)^l = x^{kl}$$

$$(xy)^k = x^k \cdot y^k$$

Some rules of exponents

$$x^1 = x$$

$$x^0 = 1$$

$$x^k + x^l = x^{k+l}$$

$$(x^k)^l = x^{kl}$$

$$(xy)^k = x^k \cdot y^k$$

More rules of exponents

$$\left(\frac{x}{y}\right)^k = \left(\frac{x^k}{y^k}\right)$$

More rules of exponents

$$\left(\frac{x}{y}\right)^k = \left(\frac{x^k}{y^k}\right)$$

$$\frac{x^k}{x^l} = x^{k-l}$$

More rules of exponents

$$\left(\frac{x}{y}\right)^k = \left(\frac{x^k}{y^k}\right)$$

$$\frac{x^k}{x^l} = x^{k-l}$$

$$x^{1/k} = \sqrt[k]{x}$$

More rules of exponents

$$\left(\frac{x}{y}\right)^k = \left(\frac{x^k}{y^k}\right)$$

$$\frac{x^k}{x^l} = x^{k-l}$$

$$x^{1/k} = \sqrt[k]{x}$$

$$x^{-k} = \frac{1}{x^k}$$

More rules of exponents

$$\left(\frac{x}{y}\right)^k = \left(\frac{x^k}{y^k}\right)$$

$$\frac{x^k}{x^l} = x^{k-l}$$

$$x^{1/k} = \sqrt[k]{x}$$

$$x^{-k} = \frac{1}{x^k}$$

Exponents in R

```r
x<-4
x^3
```

```
## [1] 64
```

Exponents in R

```r
x<-4
x^3
```

```
## [1] 64
```

```r
x^(2+3)
```

```
## [1] 1024
```

Exponents in R

```
x<-4
x^3
```

```
## [1] 64
```

```
x^(2+3)
```

```
## [1] 1024
```

```
## for base e, use exp()
exp(4)
```

```
## [1] 54.59815
```

A logarithm is the power x required to raise a base c to a given number a.

Logarithms

A logarithm is the power x required to raise a base c to a given number a.

$$\log_c(a) = x$$

$$c^x = a$$

A logarithm is the power x required to raise a base c to a given number a.

$$\log_c(a) = x$$

$$c^x = a$$

$$3^2 = 9$$

Logarithms

A logarithm is the power x required to raise a base c to a given number a.

$$\log_c(a) = x$$

$$c^x = a$$

$$3^2 = 9$$

$$\log_3(9) = 2$$

Common logarithms

- The most common log bases are 2, 10, and $e = 2.718$
- Log with base e is called a natural log, *ln*
- The R function log() has a default base e
- We use log base e to model many exponential growth processes

Logarithms are the inverse of exponents

$$10^2 = 100$$

Logarithms are the inverse of exponents

$$10^2 = 100$$

$$\log_{10}(100) = 2$$

Logarithms are the inverse of exponents

$$10^2 = 100$$

$$\log_{10}(100) = 2$$

$$e^2 = 7.389056$$

Logarithms are the inverse of exponents

$$10^2 = 100$$

$$\log_{10}(100) = 2$$

$$e^2 = 7.389056$$

$$\log_e(7.389056) = 2$$

Logarithm rules

$$\log(x \cdot y) = \log(x) + \log(y)$$

Logarithm rules

$$\log(x \cdot y) = \log(x) + \log(y)$$

$$log(x^n) = n \log(x)$$

Logarithm rules

$$\log(x \cdot y) = \log(x) + \log(y)$$

$$log(x^n) = n \log(x)$$

$$\log\left(\frac{x}{y}\right) = \log(x) - \log(y)$$

Logarithms in R

For statistics, it is safe to assume that *log* means *ln*. In R, this is the default

```
log(10)
```

```
## [1] 2.302585
```

```
log(10, base = 10)
```

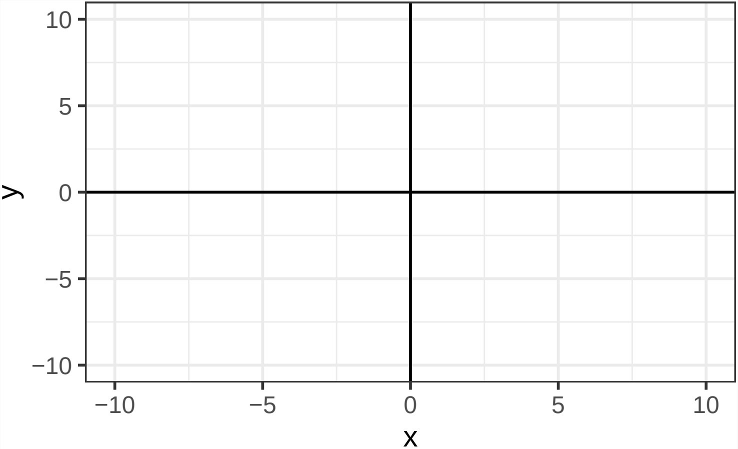
```
## [1] 1
```

Exercises

Compute the following in R. Assume $x = 6$

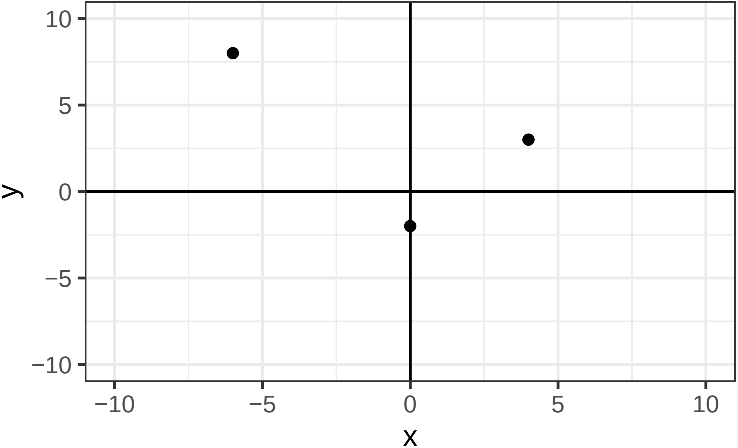
1. x^4
2. $2x^{5+x}$
3. e^x
4. $log(x)$
5. $log(2x + 3)$
6. $log(\frac{1}{2x})$

Coordinates and lines


10
5
0
−5
−10
y
−10
−5
0
5
10
x

For coordinate pairs $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, we can plot each point along an x and y axis.

Example: $(0, -2), (4, 3), -6, 8)$



The typical equation for a line is $y = mx + b$ where m is the slope and by is the y-intercept.

Lines

The typical equation for a line is $y = mx + b$ where m is the slope and by is the y-intercept.

You will often see a line expressed as a regression equation:

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

where β_0 is the y-intercept and β_1 is the slope.

The slope

Slope measures the steepness of a line. A line with a positive slope has increasing values of y as x increases. A line with a negative slope has decreasing values of y for increasing values of x.

The slope

Slope measures the steepness of a line. A line with a positive slope has increasing values of y as x increases. A line with a negative slope has decreasing values of y for increasing values of x.

We can calculate the slope with two coordinates on the line $(x_1, y_1), (x_2, y_2)$

The slope

Slope measures the steepness of a line. A line with a positive slope has increasing values of y as x increases. A line with a negative slope has decreasing values of y for increasing values of x.

We can calculate the slope with two coordinates on the line $(x_1, y_1), (x_2, y_2)$

The slope is the ratio of the difference in y values to the difference in x values.

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

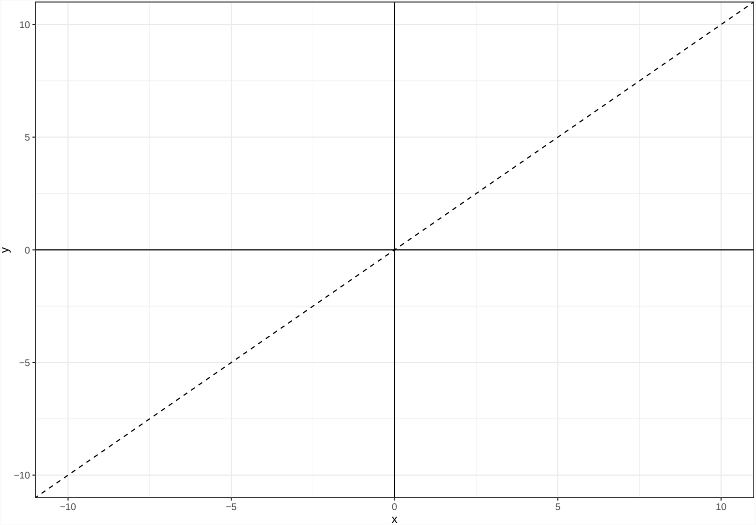
The y-intercept is the value of y when $x = 0$. If we have the value of one point on the line, and the slope, we can obtain the y-intercept

The y-intercept

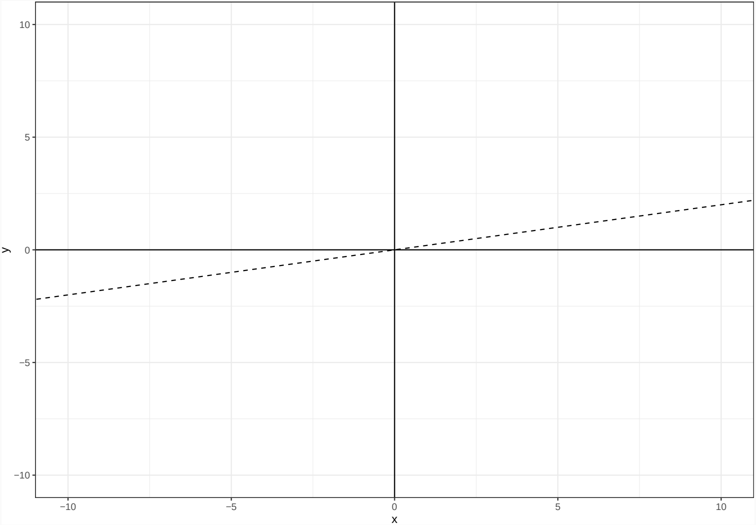
The y-intercept is the value of y when x = 0. If we have the value of one point on the line, and the slope, we can obtain the y-intercept

$$b = y_1 - m \cdot x_1$$

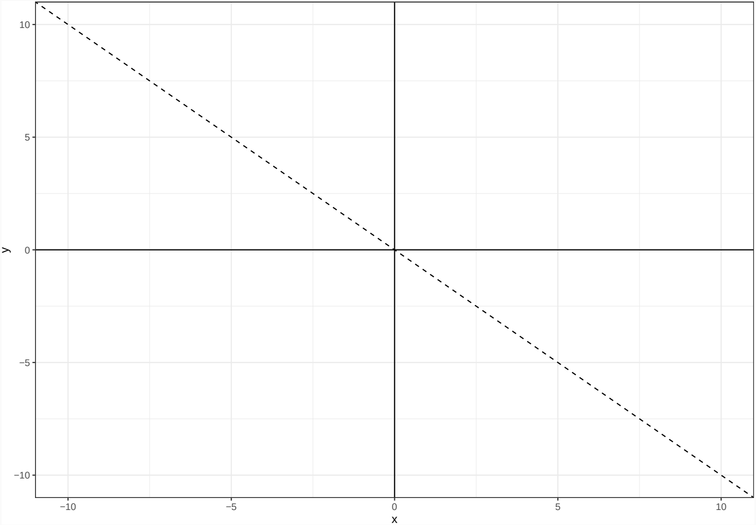
Example: intercept = 0, slope = 1



Example: intercept = 0, slope = 0.2

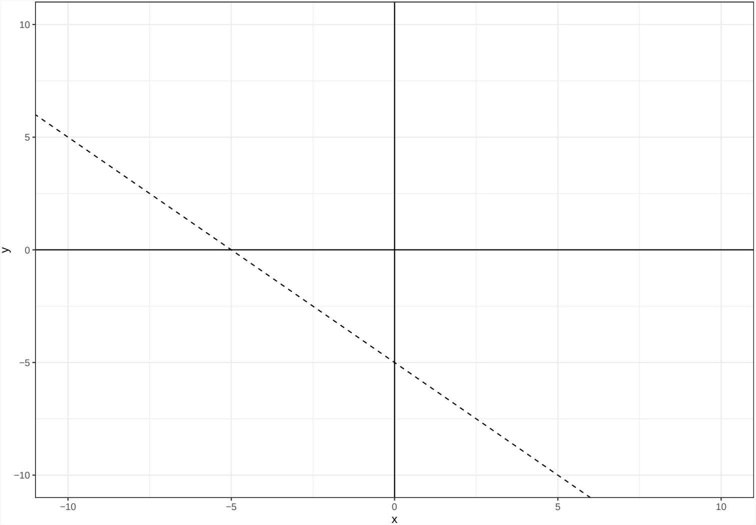


Example: intercept = 0, slope = -1

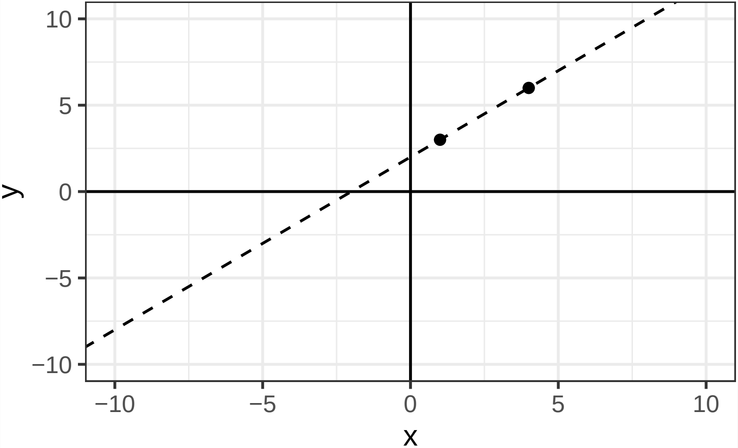


Example: intercept = 5, slope = -1

Example: intercept = -5, slope = -1



Given the points $(1, 3)$ and $(4, 6)$, the slope is $m = \frac{6-3}{4-1} = 1$ and the y-intercept is $b = 3 - 1 \cdot 1 = 2$. The equation of the line is $y = 1x + 5$



Functions

Functions, in general

A function maps each element in a set X to an element in set Y

- Linear function: $f(x) = x + 5$

Functions, in general

A function maps each element in a set X to an element in set Y

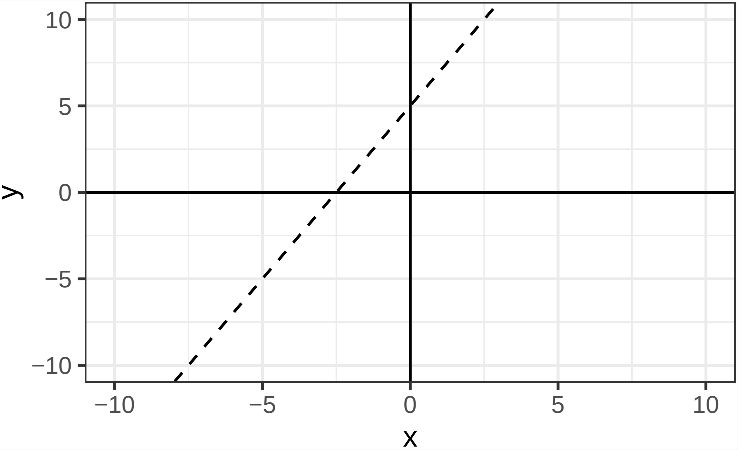
- Linear function: $f(x) = x + 5$
- Quadratic function: $f(x) = x^2 + 2x + 3$

Functions, in general

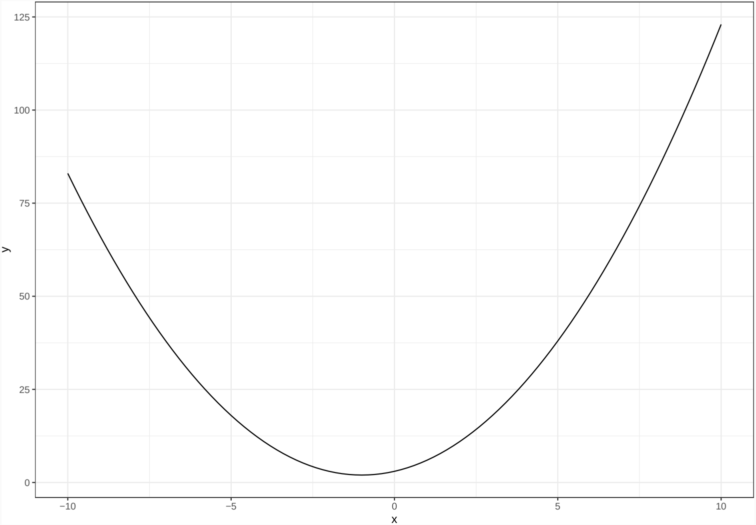
A function maps each element in a set X to an element in set Y

- Linear function: $f(x) = x + 5$
- Quadratic function: $f(x) = x^2 + 2x + 3$
- Exponential function: $f(x) = e^{2x} + 6$

$f(x) = 2x + 5$

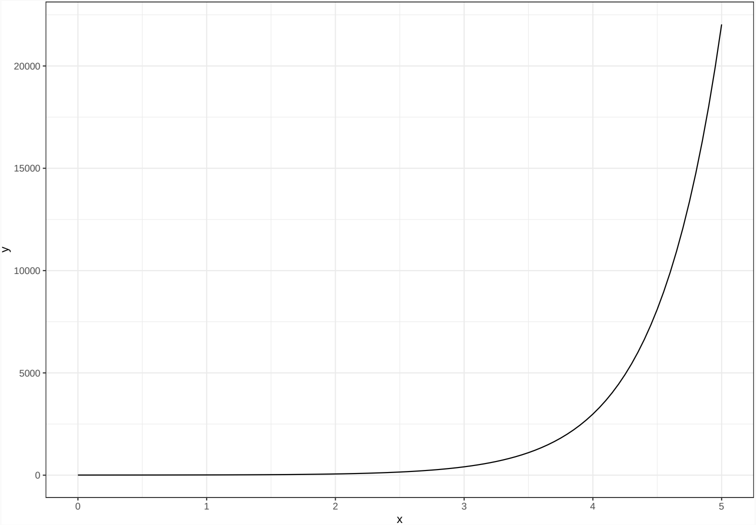


$f(x) = x^2 + 2x + 3$



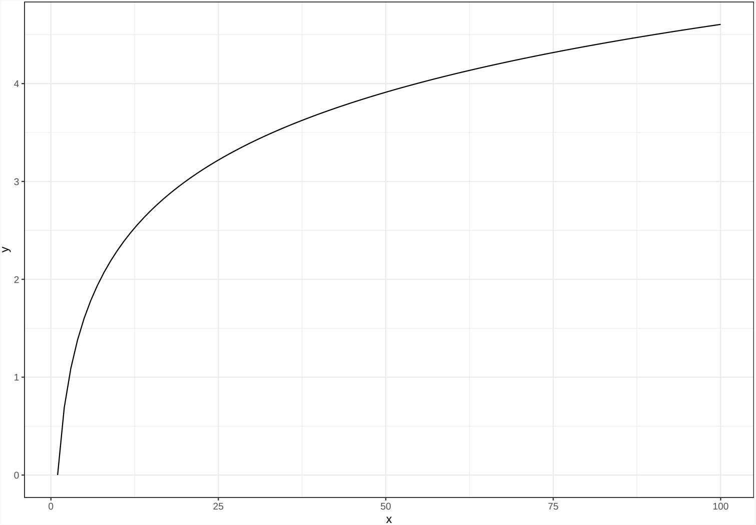
Graphical forms of functions: exponential

$f(x) = e^{2x} + 6$



Graphical forms of functions: logarithmic

$f(x) = log(x)$



Functions in R

We can easily define functions in R.

$$f(x) = 2(x+3)^3$$

Functions in R

We can easily define functions in R.

$$f(x) = 2(x+3)^3$$

```
f_x<-function(x){2 * (x + 3)^2}
f_x(3)
```

```
## [1] 72
```

Functions in R

We can easily define functions in R.

$$f(x) = 2(x + 3)^3$$

```
f_x<-function(x){2 * (x + 3)^2}
f_x(3)
```

```
## [1] 72
```

$$g(x) = \frac{x - 1}{5}$$

Functions in R

We can easily define functions in R.

$$f(x) = 2(x+3)^3$$

```
f_x<-function(x){2 * (x + 3)^2}
f_x(3)
```

```
## [1] 72
```

$$g(x) = \frac{x-1}{5}$$

```
g_x<-function(x){(x-1)/5}
g_x(4)
```

```
## [1] 0.6
```

Define and evaluate the following functions in R. Assume $x = 2$

1. $f(x) = 2x$
2. $f(x) = \frac{x}{2}$
3. $f(x) = 2(x + 1)^3$

Matrices

What is a matrix?

```
##            Murder Assault UrbanPop Rape
## Alabama      13.2     236       58 21.2
## Alaska       10.0     263       48 44.5
## Arizona       8.1     294       80 31.0
## Arkansas      8.8     190       50 19.5
## California    9.0     276       91 40.6
## Colorado      7.9     204       78 38.7
```

What is a matrix?

A matrix is a rectangular array of numbers, with dimensions expressed as rows $\times$ columns

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

So A is a 3×3 matrix, B is a 3×2 matrix, and C is a 2×3 matrix.

Matrix notation

We can identify each element of a matrix with its column and row position, where x_{ij} refers to the value in the ith row and jth column of matrix X. Note that we use uppercase letters for a matrix, and lowercase letters for elements of a matrix.

$$\mathsf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{bmatrix}$$

Vectors

Vectors are one-dimensional arrays of values. Either an n-row column or an n-column row:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$$

OR

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

A diagonal matrix has zero values except on the diagonal:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

An identity matrix is a special case of a diagonal matrix, where all values on the diagonal are equal to 1

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

These matrices are also symmetric, where all $x_{ij} = x_{ji}$. Symmetric matrices must be square.

Matrix operations

Matrix transpose

A transpose interchanges the rows and columns of a matrix, or rotates it.

The dimensions are switched, so an $n \times k$ matrix becomes a $k \times n$ matrix.

We denote a transpose with a T

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

Addition and subtraction

Two matrices (or vectors) can be added or subtracted only if they have identical dimensions. Then add or substract the correspoding elements of the matrices.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix}$$

Multplication by scalar

Matrices and vectors can be multiplied by constant values (called scalars).

$$\mathsf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} 4\mathsf{B} = \begin{bmatrix} 4 & 8 & 12 \\ 16 & 20 & 24 \end{bmatrix}$$

$$\mathsf{C} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} 3\mathsf{C} = \begin{bmatrix} 3 \\ 6 \\ 9 \\ 12 \end{bmatrix}$$