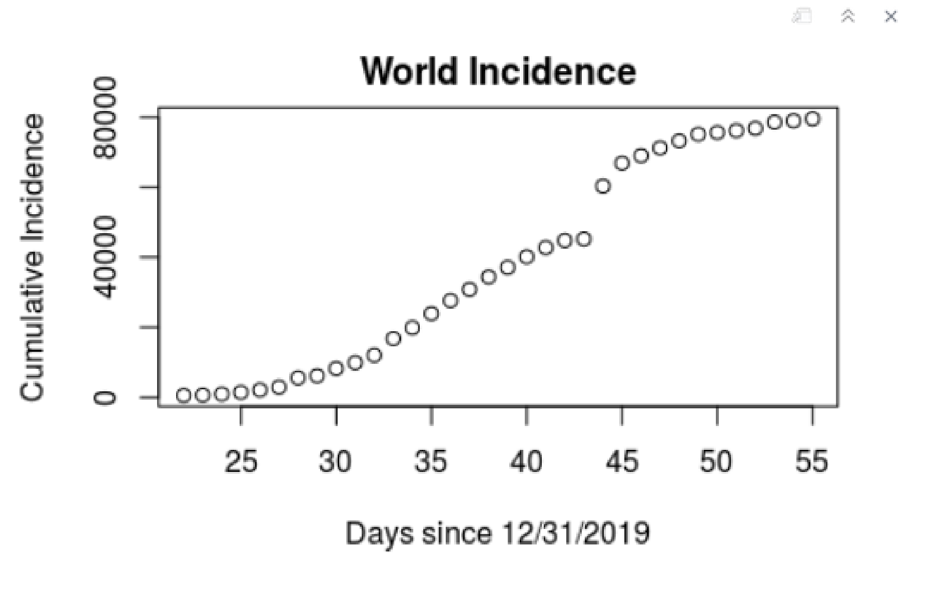
Covid and logistic curves

After completing this section, students will be able to:

- Describe the shape of a logistic curve
- Explain the meaning of the parameters in the equation $y = \dfrac{L}{1 + e^{-b(x-m)}}$ for a logistic curve.
- Use R to fit a logistic curve to data
- Predict the number of cases at future times from a logistic model.
- Predict the length of time of an epidemic from a logistic model.
- Use a logistic model for one country, and proportional reasoning, to estimate reasonable parameters for a logistic curve for another country.

Logistic curves

If we plot the entire cumulative incidence curve for the world through Feb 24, it is clearly not exponential. The shape of this curve is called *logistic*.



One equation for a logistic curve is given by:

$$y = \frac{L}{1 + e^{-b(x-m)}}$$

This can also be written as:

$$y = \frac{L}{1 + exp(-b(x - m))}$$

where L, b, and m are *parameters* that you need to fill in with numbers.

Go to desmos.com and in the graphing calculator, type:

$$y = 10/(1 + \exp(-2(x - 3)))$$

Describe the shape of the curve. Where does it have:

- horizontal asymptote(s)?

  Note: A horizontal asymptote is a horizontal line where the curve levels off as the x-values get very big, or very small

- an inflection point?

  Note: An inflection point is a point at which the curve changes from concave up (like a bowl that holds water) to concave down (like an upside down bowl that spills water)

Next, enter into Desmos the equation

$$y = L/(1 + \exp(-b(x - m)))$$

Desmos will ask you if you want to create a slider for the variables L, b, and m. Click "all".

You can now use the sliders to change the values of L, b, and m from their default values of 1, 1, and 1.

By experimenting, and looking at the resulting graph of this curve, and see if you can tell what information about the shape of the curve is given by the parameters:

- L?

- m?

- b?

If we can fit a logistic curve to early data in an epidemic, then we can use the parameters to make predictions.

- Which parameter(s) can be used to figure out how many people, total, will ultimately get infection (if the epidemic continues to run its course according to this same logistic curve)?

- Which parameter(s) can be used to figure out when we will have the most daily new cases and hospitals will be the most full?

- Which parameter(s) can be used to figure out how much longer the pandemic will last (again, assuming no changes from the logistic curve)?

Fitting logistic curves

Fitting a logistic curve to a data set in Python is similar to fitting a line. The key steps are

- Define the logistic curve

  ```
  def logistic(x, L, b, m):
    return L/(1 + np.exp(-b*(x - m)))
  ```

- Use the curve_fit function from the scipy.optimize to fit the curve.

  ```
  (L1, b1, m1), pcov = opt.curve_fit(logistic, FILL_IN_XVALUES, FILL_IN_YVALUES)
  ```

- If you get an error message with an "OptimizeWarning", and weird parameters as your answer, then retry using bounds . The bounds give you the lower end and the upper end of reasonable values for the parameters.

  ```
  (L2, b2, m2), pcov = opt.curve_fit(logistic, df0.Days[0:100], df0.US[0:100],
  bounds = ([1500000, 0, 100] , [4000000, 5, 150]))
  ```

The file ”logisticGrowthChinaDataPython.ipynb” in tinyurl.com/math115unc > Week 13 goes through an example of fitting a logistic curve to cumulative incidence for China for the first 100 days.

Suppose we had tried to predict the future, at day 45 (Feb 14), by fitting a logistic curve to only the first 45 days of data in China.

What would we have predicted for the total number of cases in China, longterm?

What would we have predicted for the amount of time the epidemic would last in China?

How far off would we have been?

Example. Fit a logistic curve to the US data for the first 150 days.

Based on this curve, what would we predict for how many people total would be invected in the US?

Based on this curve, what would we predict for how long until the epidemic is over in the US?

Why are these estimates so far off?

Example. Fit a logistic curve to the first 100 days of 2020 in Italy.

How well does it appear to fit the data from the first 100 days? Comment on the fit – does it look good, so-so, or bad?

If we were at day 100 and using this model, what would we predict for the total number of people in Italy who would ultimately get covid?

What would we predict for how long the epidemic would last in Italy?

If we extended it through the first 150 days, would it continue to be a good fit?

Now fit a logistic curve to the first 150 days of 2020 in Italy.

Does the Italy data appear to be exactly logistic in shape? Why might it deviate from logistic?

Example. Fit a logistic curve to cumulative incidence data from South Korea for days 0 - 72.

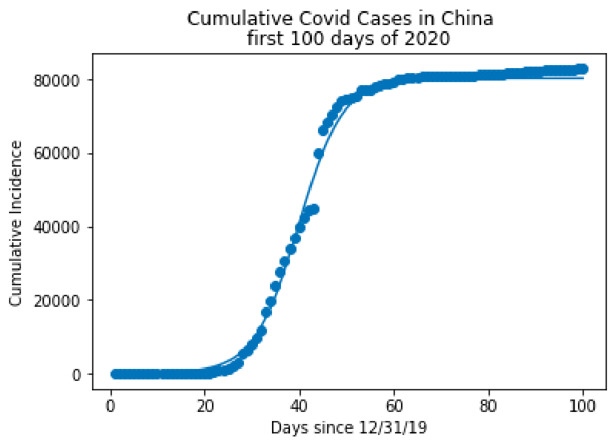
Can the logistic fit curve be used to estimate the basic reproductive number r_0 for a given country?

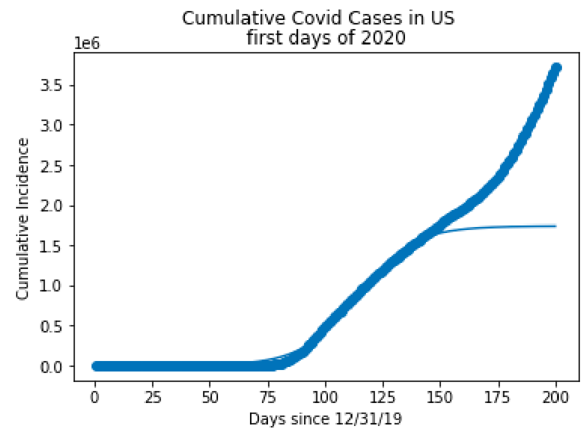
Example. Hubei province

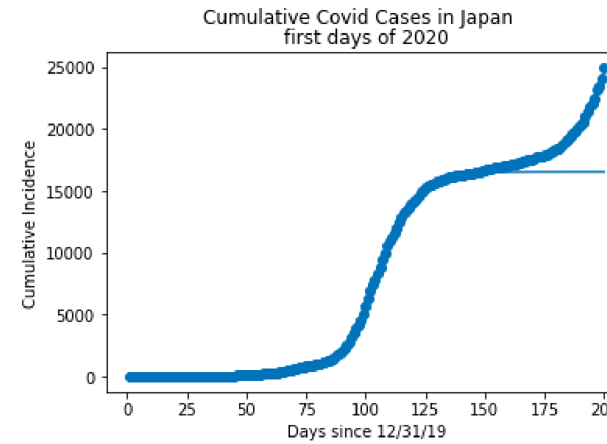
- Fit a logistic curve to Hubei province in China.

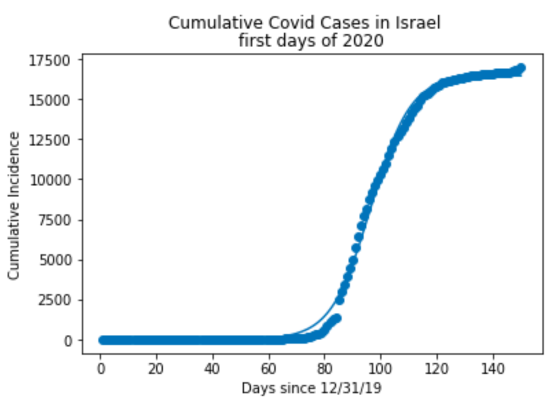
- Try to approximately reproduce this curve in the SIR model using different values of r_0.
- What is your estimate of r_0 in Hubei province?

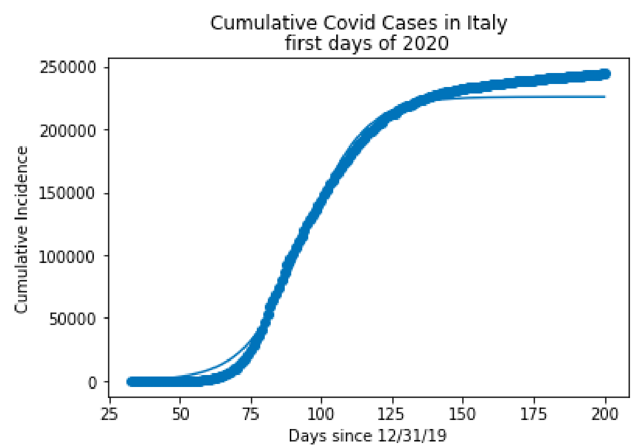
How is the total number of cases different in the SIR model from the fit curve in the actual data. Can you give possible explanations for the differences?


Cumulative Covid Cases in China
first 100 days of 2020
Cumulative Incidence
Days since 12/31/19


Cumulative Covid Cases in US
first days of 2020
1e6
Cumulative Incidence
Days since 12/31/19


Cumulative Covid Cases in Japan
first days of 2020
Cumulative Incidence
Days since 12/31/19


Cumulative Covid Cases in Israel
first days of 2020
Cumulative Incidence
Days since 12/31/19


Cumulative Covid Cases in Italy
first days of 2020
Cumulative Incidence
Days since 12/31/19