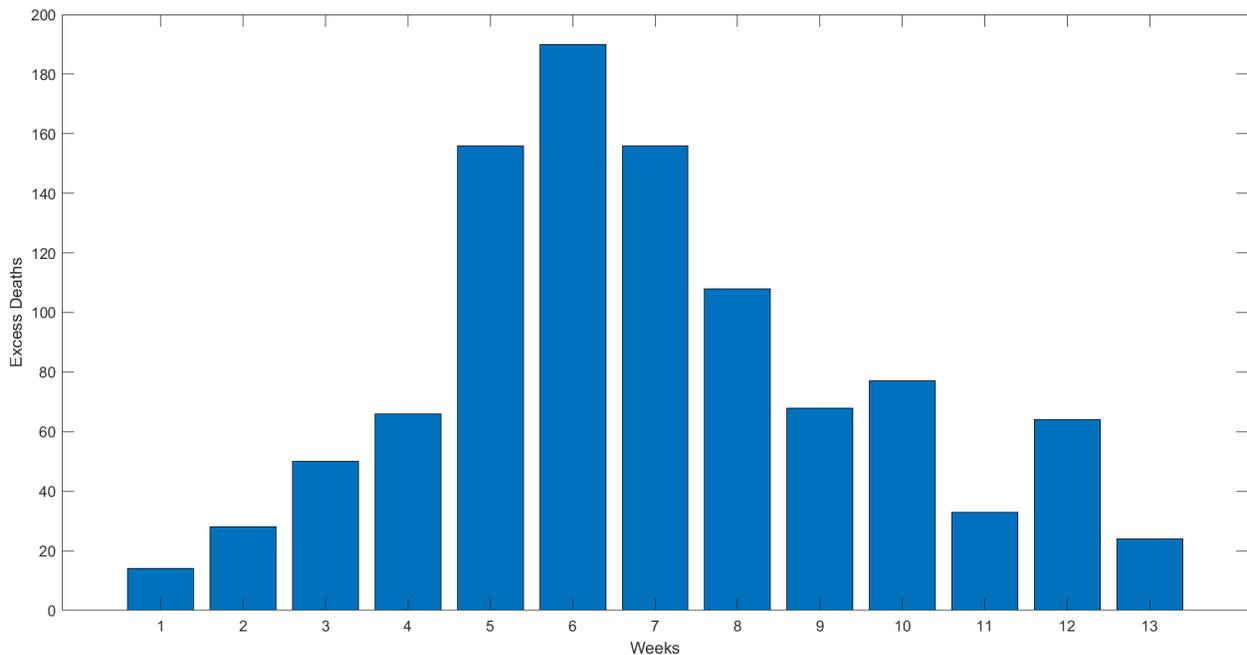


What makes a virus an epidemic?

Background During the winter of 1968-69, the United States was swept by a virulent new strain of influenza named *Hong Kong flu*. At that time, no flu vaccine was available, so many more people were infected than would be the case today. We will study the spread of the disease through a single urban population, New York City. The data displayed in the following table are weekly "excess" pneumonia-influenza deaths, that is, the numbers of such deaths in excess of the average numbers to be expected from other sources. The graph displays the same data (Source: Centers for Disease Control)

Week	Flu-related deaths	Week	Flu-related deaths
1	14	8	108
2	28	9	68
3	50	10	77
4	66	11	33
5	156	12	65
6	190	13	24
7	156		



Submit documentation of your answers to the following questions.

1. Create a linear and logarithmic plot of flu related (weekly) deaths. What observations can you make about the trend? What is the maximum number of weekly deaths? When did the maximum occur?
2. Relatively few flu sufferers die from the disease or its complications, even without a vaccine. However, we may reasonably assume the number of excess deaths in a week was proportional to the number of new cases of flu in some earlier week. Thus, the figures reflect (proportionally) the rise and subsequent decline in the number of new cases of *Hong Kong flu*. At any given time during a flu epidemic, we want to know the number of people who are infected. We also want to know the number who have been infected and have recovered, because we typically assume that they are now immune to the disease. (To simplify our model we will include the dead amongst the recovered. We will also ignore the people who are born or die from non-flue cases or move in and out of the city during those weeks.) The remainder of the population is still susceptible. With these assumptions, the fixed total population (approximately 7,900,000) may be divided into three categories,
 - those who are infected, $I(t)$
 - those who have recovered, $R(t)$ and,
 - those who are still susceptible, $S(t)$,

such that the independent variable t is time. We can normalize each of those functions by creating a second set of equations that represent the fraction of the total population in each category. If N is the total population, we have:

- $i(t) = I(t)/N$
- $r(t) = R(t)/N$
- $s(t) = S(t)/N$

Under the assumptions we have made:

- a. How do you think $s(t)$ should vary with time? How should $r(t)$ vary with time? How do you think $i(t)$ should vary with time?
 - b. Sketch what you think the graph of each of these functions looks like
 - c. Explain why, at each time t , $s(t) + i(t) + r(t) = 1$
3. Next, let's make some assumptions about the rates of change of our dependent variables. No one is added to the susceptible group, since we ignore birth and migration. The only way an individual leaves the susceptible group is by becoming infected. The number of susceptibles depends on the number already susceptible, the number of individuals already infected and the contact between the two groups. Let us suppose that the coefficient β represents the contacts per day between the infected and the susceptible that are sufficient to spread the disease. If we assume homogeneous mixing of the populations, each infected individual generates $\beta * s(t)$ new infected individuals per day.

We also assume that a fixed fraction μ of the infected group will recover during any given day. For example, if the average duration of infection is three days, then, on average, one-third of the currently infected population (actually "infectious" population recovers each day) recovers each day. We will ignore the "recovered" (non-infectious) people, for now, who still

feel miserable, and might even die later from pneumonia. As a result we have the following differential equations

- The Susceptible Equation: $ds/dt = -\beta * s(t) * i(t)$
- The Recovered Equation: $dr/dt = \mu * i(t)$
- The Infected Equation: $di/dt = \beta * s(t) * i(t) - \mu * i(t)$

- a. Explain why $ds/dt + di/dt + dr/dt = 0$?
- b. Describe both components of the infected equation.
- c. What do the minus signs represent?

4. Finally we complete our model by giving each differential equation an initial condition. For this particular virus hardly any was immune, so almost everyone was susceptible. That would suggest $s(0) = 1$. We will assume there was a trace level of infection in the population, say, 10 people. Then $i(0) = 0$ and $r(0) = 0$. We don't know the values for the parameters β and μ . We can estimate them, and then adjust them as necessary to fit the excess death day. We estimated the average period of infectiousness as three days. That would suggest $\mu = 1/3$. If we guessed that each infected would make a possibly infected contact every two days, then $\beta = 1/2$.

The following program simulates the effects of the virus on the populations of New York City.

```
function dpop = SIRmodel(pop,b,k) % this is the function header, note time is not
                                % a variable
s = pop(1); % pop(1) is the input susceptible population. pop is a vector.
i = pop(2); % pop(2) is the input infected population
r = pop(3); % pop(3) is the input recovered population

dpop=[ % dpop is the rate of change of each of the population variables
      - b*s*i; % this the equation for the rate of change of s, b is for beta
      (b*s*i - k*i); % this the equation for the rate of change of i, k is for mu
      k*i]; % this the equation for the rate of change of r
end
```

- a. Type the program into an m-file and save the function. Run the function with the givens provided.

```
pop = [ 1 - (10/9700000),10/9700000,0]
```

```
μ = 1/3
```

```
β = 1/2
```

```
SIRmodel(pop, μ, β)
```

What results do you get for dpop? Explain the significance of positive and negative signs.

- b. Euler's method is a numerical solution method for a system of differential equations. if we know \mathbf{y}_0 (i.e. $\mathbf{i}(0)$, $\mathbf{r}(0)$, and $\mathbf{s}(0)$) and we have a way to calculate $\mathbf{dy/dt}$ at any point (\mathbf{t}, \mathbf{y}) (i.e. dpop), then it follows that we can calculate a sequence of \mathbf{y} -values from the "slope and intercept equation",

$$\mathbf{y}_n = \mathbf{y}_{n-1} + (\mathbf{dy/dt})_{n-1} * \Delta t$$

where Δt is a small change in time. For the SIR model, we want the dependent variables to be \mathbf{s} , \mathbf{i} , and \mathbf{r} . Use the following script to apply Euler's method to the SIRmodel function. Display a graphical solution of \mathbf{s} , \mathbf{i} , and \mathbf{r} (POP(1,:), POP(2,:), POP(3,:)) on a single plot.

```
function dpop = SIRmodel(t, pop,b,k) % this function now has t as a variable
s = pop(1);      % pop(1) is the input susceptible population. pop is a vector.
i = pop(2);      % pop(2) is the input infected population
r = pop(3);      % pop(3) is the input recovered population

dpop=[           % dpop is the rate of change of each of the population variables
    - b*s*i;      % this the equation for the rate of change of s
    (b*s*i - k*i); % this the equation for the rate of change of i
    k*i];        % this the equation for the rate of change of r
end

%Euler's Method

b = (1/2);
```

```

u = (1/3);

dt = 10;                % let  $\Delta t = 10$  days

tspan = 0:dt:140;      % let t range up to 150 days

pop0 = [1-1.27e-6;1.27e-6;0];    % this is [s(0),i(0),r(0)]

POP(:,1) = pop0;       % POP is the matrix of population data

for i = 1:(150/dt)     % forward euler in a for loop

    POP(:,i+1) = POP(:,i) + dt * SIRmodel(tspan, POP(:,i),b,u);

end

```

- c. Change Δt to 1 day and replot the solutions. What observations do you have in comparison to the previous solution?
- d. *ode45* is a built-in ordinary differential equation solver. Implement the following code and replot the solutions. What observations do you have in comparison to the previous solution?

```

function dpop = SIRmodel(t, pop,b,k) % this function now has t as a variable

s = pop(1);    % pop(1) is the input susceptible population. pop is a vector.
i = pop(2);    % pop(2) is the input infected population
r = pop(3);    % pop(3) is the input recovered population

dpop=[         % dpop is the rate of change of each of the population variables
    - b*s*i;   % this the equation for the rate of change of s
    (b*s*i - k*i); % this the equation for the rate of change of i
    k*i];     % this the equation for the rate of change of r

end

```

```
[t,pop] = ode45(@(t,pop)SIRmodel(t,pop,b,k),tspan,pop0);
```

```
plot(tspan,pop(:,1),tspan,pop(:,2),tspan,pop(:,3))
```

```
xlabel('Time')
```

```
ylabel('Population')
```

```
title('ode45')
```

I've posted an alternate script for this approach in Canvas-[funhandleapproach.m](#)

- e. Take the output of the infected population and find its maximum over the region of interest. Plot the maximum (**as a separate marker**) on the previous figure.
5. While the value of μ is related to the flu's infectious period the value of β is approximation of the flu's "infectiousness", and there is no way to directly observe β . Taken together, these values determine the disease's spread. In reality, β and μ are a function of a city's population, density, and "mixing rate".
- a. Let's experiment with changes in β . Keep μ fixed at $\frac{1}{3}$. Create a function (with a loop) and plot $i(t)$ with different values of β . Use equally spaced values, ranging from 0.5 to 2.0. You will also want to increase the interval of interest for time to see the results for the lower levels of contact rate. Store the peak levels of the infected population in an array and when they occur. Then plot those results versus β . Describe how these changes affect the graph of $i(t)$ in the context of the SIR model.
 - b. Let's experiment with a change in μ . Return β to $\frac{1}{2}$. Create a function (with a loop) and plot $i(t)$ with different values of μ . Use equally spaced values, ranging from 0.1 to 0.6. You will also want to increase the interval of interest for time to see the results for the lower levels of contact rate. Store the peak levels of the infected population in an array and when they occur. Then plot those results versus μ . Describe how these changes affect the graph of $i(t)$ in the context of the SIR model.
 - c. Overlay an appropriately scaled plot of flu-related deaths (**our original data set with the deaths divided by 7900000**) with your model of the infected population ($i(t)$). Which model parameters (β, μ) in the given range seem reasonable? Explain your conclusion.
6. The ratio of β to μ , is known as the contact number, $R_o = \beta/\mu$. The contact number is a combined characteristic of the population and the disease. In similar populations, it measures the relative contagiousness of the disease, because it tells us indirectly how many

of the contacts are close enough to actually spread the disease. This formalism (**c**) allows for the derivation of the following expression from di/dt divided by ds/dt ,

$$i = -s + (1/R_0) \ln(s) + i(0)$$

which is a time-independent equation. There are two times when we know (or can estimate) the values of i and s : at $t = 0$ and $t = \infty$. $i(0)$ is approximately 0. Thus:

$$i = -s + (1/R_0) \ln(s)$$

After a long time, $i(\infty)$ is approximately 0 again, and $s(\infty)$ is some constant. If there has been good reporting of the numbers who have contracted the disease, then

$$R_0 = (\ln s_{inf}) / (s_{inf} - 1)$$

Use a numerical solution from Part 5, part c to estimate the value of $s(\infty)$. Use this value to calculate the contact number R_0 for the Hong Kong flu. Compare your calculated value with the one you get by direct calculation from the definition, $R_0 = \beta/\mu$.

7. When experimenting with the relative sizes of β and μ in Part 5, you found that if β is small enough relative to μ , then no epidemic can develop. In Part 6, if the contact number R_0 is small enough, then there will be no epidemic. But another way to prevent an epidemic is to create *herd immunity* by reducing the initial susceptible population (s_0) artificially through inoculation.

The point of inoculation is to create herd immunity by stimulating in as many people as possible the antibodies that confer immunity without actually giving those people the disease. The fraction of the population that must be inoculated to obtain herd immunity will depend upon the contact number. Mathematically speaking, if $i'(0) = (\mu) (R_0 s_0 - 1) i_0$, then as long as s_0 is less than $1/R_0$, the rate of change in the infectious population should be negative; and an epidemic will not occur.

- a. From 1912 to 1928, the contact number for measles in the U.S. was 12.8. If we assume that R_0 is still 12.8 and that inoculation is 100% effective -- everyone inoculated obtains immunity from the disease -- what fraction of the population must be inoculated to prevent an epidemic?
- b. Suppose the vaccine is only 95% effective. What fraction of the population would have to be inoculated to prevent a measles epidemic?

References:

1. David Smith and Lang Moore, "The SIR Model for Spread of Disease - Herd Immunity," *Convergence* (December 2004)
2. Gon, B. (2020, April 26). Epidemic Modeling 101: Or why your CoVID19 exponential fits are wrong. Retrieved from <https://medium.com/data-for-science/epidemic-modeling-101-or-why-your-covid19-exponential-fits-are-wrong-97aa50c55f8>