Make Your Own SIR Model

pandemsim.com/data/index.php/make-your-own-sir-model/

Modeling Infectious Disease Spread with Excel

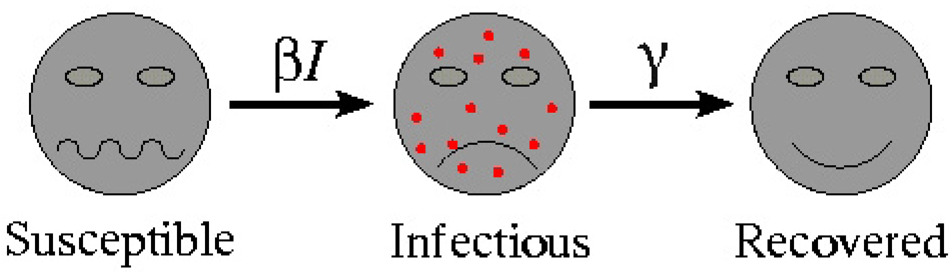
Medical researchers and mathematicians have developed a series of sophisticated mathematical models to describe the spread of infectious diseases. But even a simple model is useful to predict how long an outbreak of a disease, for example the flu, will last and how many people will be sickened by it.

The oldest and most common model is the SIR model which considers every person in a population to be in one of three conditions:

- S = Susceptible to becoming infected
- I = Infected through contact with someone already infected
- R = Recovered, no longer sick or infected.

Through time a person may move from being susceptible to infected to recovered, so the number of people in each condition changes, but the total of S + I + R is constant. S + I + R = N where N represents the entire population and is identified as closed system population.

This is a compartmental model, with S, I and R being compartments. Every person starts off in a compartment and many move to others over time. Graphically the compartment model looks like the figure below with the rates of movement between compartments given as Greek letters above the arrows indicating direction of movement.



Movement rates between classes of the SIR model

(from Keeling, 2001)

Assumptions & Parameters

This is a steady-state model with no one dying or being born, to change the total number of people. In this model once someone recovers they are immune and can't be infected again. More sophisticated models allow re-infections. The model also assumes that a disease is passed from person to person. The SIR model can't be used for diseases that spread other ways, such as by insect bites.

To run this model, you need to know the following:

- initial population, S (initial number of people who are susceptible),
- initial number of infected people, I
- Infection rate,

  ß (Greek letter beta, the rate (#/day) that susceptible people become infected),
- recovery rate, γ (Greek letter gamma, the rate that infected people recover),
- time increment, T (the time interval or steps during which changes occur.) T is usually set at one day and because its value is 1, is often ignored. For rapidly spreading outbreaks T might be one hour or some other short interval.

How do you know the values of these parameters? The number of susceptible people (S) can be the population of a city or town where the outbreak occurs. The number of initially infected people (I) is a guess unless it is known, for example, that a single traveller brought the disease into a community. The infection rate (

ß) and the recovery rate (γ) can be selected from rates determined from prior outbreaks, but they often vary for different outbreaks of the same disease.

Equations

With these parameters, the number of people at any time who are Susceptible, Infected, or Recovered can be calculated with these equations:

$S_n = S_{n-1} – ((S_{n-1}/S) * ($

$\beta * I_{n-1}))$

$I_n = I_{n-1} + (S_{n-1}/S) *$

$(\beta * I_{n-1}) – (I_{n-1} * \gamma)$

$R_n = R_{n-1} + (I_{n-1} * \gamma)$

Because of subscripts and Greek letters these equations look complicated, but they aren't really!

These equations calculate the number of people in each condition today (n), based on the number yesterday (n-1) and the rates of change,

ß and γ. The subscript n means the number in one time interval, and n-1 means the number in the previous interval. So with a time interval of one day, then the first equation:

$S_n = S_{n-1} – ((S_{n-1}/S) * ($

$ß * I_{n-1}))$

can be understood as:

The number of <u>susceptible</u> *people today (S_n) equals the number yesterday (S_{n-1}), MINUS the percentage of people who become infected today (which is yesterday's number of susceptible people (S_{n-1}) divided by the original number (/S)), times their rate of infection (*

ß), times how many people were infected (I_{n-1}) yesterday.

The number of susceptible people today equals the number who were susceptible yesterday minus the number who become infected today. As long as the disease is spreading, the number not yet infected – the remaining susceptibles – decreases every day.

The number who become infected today equals yesterday's number of susceptibles times the rate of infection, but it may seem odd that we also multiply that result by how many were <u>infected</u> yesterday. The reason is that *the rate of infection is for each infected person*. If 3 people are infected the chance of anyone else becoming infected is 3 times as high as when only 1 person is infected. So any estimate of the rate of spread of the disease requires knowledge of the infection rate and the numbers of initially infected and initially susceptible people.

<u>The 2nd equation says:</u>

$I_n = I_{n-1} + (S_{n-1}/S) * ($

$ß * I_{n-1}) – (I_{n-1} * γ)$

The number of <u>infected</u> *people today* (I_n) *equals the number who were infected yesterday* (I_{n-1}), *PLUS the number of susceptible people who became infected today, MINUS the number of infected yesterday who recovered.*

At the beginning of an outbreak the number of people getting infected every day is probably larger than the number recovering, so the number of infected will keep rising until more people recover than get infected.

<u>The 3rd equation says:</u>

$R_n = R_{n-1} + (I_{n-1} * γ)$

The number of <u>recovered</u> *people today* (R_n) *equals the previous number who had recovered, PLUS the number who of infected people yesterday who recovered today.*

The number of susceptible people always decreases, but the number of infected and recovered initially rise and then decline, as people get sick and then get better.

Let's make a simple calculation with these realistic values for a <u>flu</u> outbreak:

initial susceptible population, S = 1000 initial infected people, I = 1

rate of infection,

$\beta = 0.29$/day

rate of recovery, $\gamma = 0.15$/day

time increment, T = 1 day

Note that the transition rates tell you how many days it takes to double the number of people who are infected or recovered.

Since

$\beta = 0.29$/day, the time to double the number of infected people is about 1/0.29 = 3.4 days. And $\gamma = 0.15$ implies that it takes 1/0.15 = 6.7 days to recover. Infections quickly outnumber recoveries.

To start your model you set the number of infected people for the previous day, $I_{n-1} = 1$, and the number of recovered $R_{n-1} = 0$.

Now you calculate numbers of people in each compartment at the end of day 1:

$S_n = S_{n-1} - ((S_{n-1}/S) * ($

$\beta * I_{n-1}))$

$S_1 = 999-((999/1000) * (0.29 * 1)) = 998.7$ are susceptible

$I_1 = I_{n-1} + (S_{n-1}/S$

$* \beta * I_{n-1}) - (I_{n-1} * \gamma)$

$= 1 + (999/1000) * (0.29 * 1) - (1 * 0.15) = 1 + 0.29 - 0.15 = 1.14$ people are infected

$R_1 = R_{n-1} + I_{n-1} * \gamma$

$= 0 \quad + 1 * 0.15 = 0 + 0.15 = 0.15$ people have recovered

For Day 2 (n = 2):

$S_n = S_{n-1} - ((S_{n-1}/S) * ($

$\beta * I_{n-1}))$

$S_2 = 998.7 - ((998.7/1000) * (0.29 * 1.14)) = 998.4$ are susceptible

$I_2 = I_{n-1} + (S_{n-1}/S *$

$ß) – (I_{n-1} * γ)$

$= 1.14 \quad + (998.7/1000 * 0.29) – (1 * 0.15) = 1.14 + .29 – 0.15 = 1.30$ people infected

$R_1 = R_{n-1} + I_{n-1} * γ$

$= 0.15 \quad + 1 * 0.15 = 0.15 + 0.15 = 0.3$ people recovered.

The decrease in numbers of susceptibles, and the increase in numbers of infected and recovered, are all very small on day 1 and day 2. This shows that the disease is not spreading rapidly. But when you plot these data over the course of the outbreak you may be surprised!

Let Excel Do it!

These equations can be calculated by hand for each day of an outbreak, but it's a lot easier to make an Excel spreadsheet to do it. So make one!

1. Make a small table of variables at the top of your spreadsheet so you can easily change a variable to see the effect on your model – see A5 to B12 in the Excel image below. Use these realistic values for the flu outbreak that you previously hand-calculated the first two days results:

N = 1000 people

I = 1 person; so S = 1000-1 = 999

ß = 0.29/day

γ = 0.15/day

2. Here are the equations used for Line 6 (of the Model Output figure below) and all following lines:

T: =D5+1

S: =E5-((E5/B$5)*(B$8*F5))

I: =F5+(E5/B$5)*(B$8*F5)-(F5*B$9)

R: =G5+(F5*B$9)

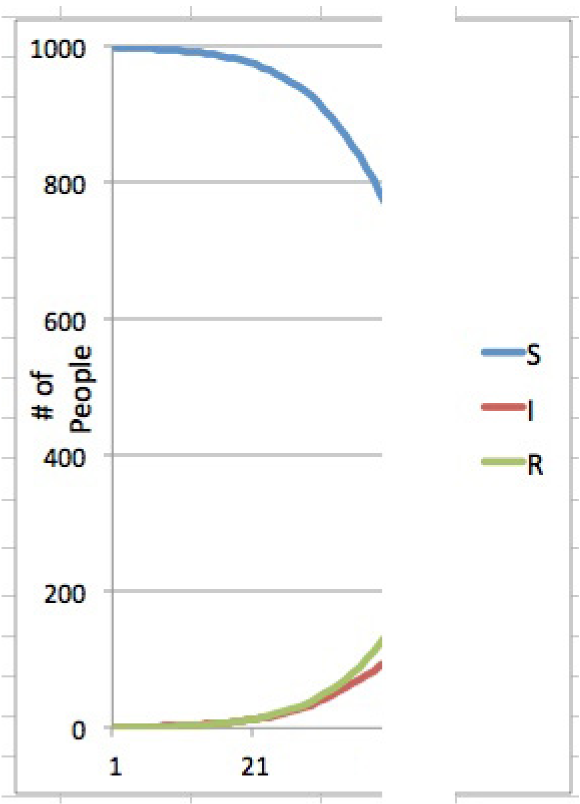
Why are these equations not used for line D5 to H5?

(Line D5 to H5 give the initial conditions)

3. Here is the spreadsheet including the output for the first 14 days of the outbreak. Your calculation should extend for 100 days to see the full development of the outbreak.

	A	B	C	D	E	F	G	H
1	**SIR Modeling for Flu**							
2	following Tassier (2013): *The Economics of Epidemics*							
3					Model Output			
4	Definitions	Values		T	S	I	R	N
5	S = Susceptible; initial	999		1	999.00	1.00	0.00	1000
6	I = Infected; initial	1		2	998.71	1.14	0.15	1000
7	R = Recovered; initial	0		3	998.38	1.30	0.32	1000
8	ß = fraction of susceptibles infected/day	0.29		4	998.00	1.48	0.52	1000
9	γ = fraction of infectives who recover/day	0.15		5	997.57	1.69	0.74	1000
10	T = time interval; usually days	1		6	997.08	1.92	0.99	1000
11	n = number of people on day n	varies		7	996.53	2.19	1.28	1000
12	N = total number of people = S+I+R	1000		8	995.89	2.50	1.61	1000
13				9	995.17	2.84	1.98	1000
14				10	994.35	3.24	2.41	1000
15	Equations			11	993.41	3.69	2.90	1000
16	$S_n = S_{n-1} - ((S_{n-1}/S) \cdot (ß \cdot I_{n-1}))$			12	992.35	4.20	3.45	1000
17	$I_n = I_{n-1} + (S_{n-1}/S) * (ß * I_{n-1}) - (I_{n-1} \cdot γ)$			13	991.14	4.78	4.08	1000
18	$R_n = R_{n-1} + (I_{n-1} \cdot γ)$			14	989.77	5.44	4.80	1000

4. Use the Charts-Line option of Excel to create a graph plotting the curves for S, I, and R over the 100 days duration of the outbreak. Here is the first 40 days, with the Y-axis being the number of people in each compartment, and the X-axis being days since the outbreak began. Watch what happens to your graph when you plot the full 100 days!



5. Write an interpretation of what the curves above tell you about the outbreak, including how long it lasts, when it reaches a maximum, and how many people become ill. What information do the slopes of the lines provide?
6. This is a very simple model. What factors does it not consider that are important? Examples include demographic factors (e.g. people of different ages may have different susceptibilities, and people are born and die every day) and preventative actions (e.g. closing schools to slow the spread). How might one such factor be mathematically added to the model?
7. Now try your SIR model with a different disease. Here are the parameters for <u>measles:</u>
   1. N = 1000 people
   2. I = 1 person; so S = 1000-1 = 999
   3. ß = 1.7/day
   4. γ = -0.14/day
8. Based on these ß and γ values how do you expect this outbreak to be different from the flu one? Does your model output confirm that?