www.simiode.org
SIMIODE
A SYSTEMIC INITIATIVE FOR MODELING INVESTIGATIONS & OPPORTUNITIES WITH DIFFERENTIAL EQUATIONS

STUDENT VERSION

Simulating the Spread of the Common Cold

Richard Corban Harwood
Department of Mathematics
George Fox University
Newberg OR USA

STATEMENT

Set-Up Instructions

A certain number (x_0) of residents on a floor of a residence hall (with a total of N residents) have the common cold and have left traces of the virus at various places. In this simulation, beans scattered onto a printed floor plan represent residents susceptible to the cold milling around this floor. Those that come in contact with the virus become infected and leave more traces of the virus. How does the number of residents infected vary with time?

1. Get into groups of four and pick up two printed floor plans (see example in Figure 1), four paper clips, and $N = 30$ beans *(optional: also pick up two plastic cups, two overhead slide pieces, and two dry erase markers)*. Choose two group members to be "Recorders" and two to be "Simulators," one for each printed floor plan.
2. For each floor plan, the Simulator should fold it in a way that lays the rooms flat with raised paper sidewalls. For the example in Figure 1, fold along the dotted lines in an alternating fashion (in-out-in) to create a 'w' shape. Then pinch the two floors together and fold in between them to make the second and third floors lay down horizontally. The Recorder should then use two paper clips to fasten the middle ridge at both ends.
3. Each Recorder sets up five tables, each with four rows, to record the results of five simulations. Label the four rows *Time* (round number, starting with zero), *Susceptible* (number dumped), *Infected* (number removed from the board and set aside), and *Change in Infected* (number infected in this round and used to get the number *Infected* in the next round).

In-Class Simulation

Each Simulator-Recorder pair will do the following to run and record data from five simulations.

1. The Recorder chooses a starting infected count, x_0, randomly chosen between 1 and $N-1$ and fills in the table for this simulation with 0 (Time), $N - x_0$ (Susceptible), x_0 (Infected). *To aid model development, start the first simulation with* $x_0 = 1$.
2. The Simulator sets x_0 beans to the side as infected and marks with an 'X' that many rooms on the second floor chosen at random locations (excluding bathrooms/hallways/stairwells).
3. The Recorder walls off the two ends of the second floor with their hands, while the Simulator shakes and dumps the susceptible beans from their hand/cup onto the center of the floor. If any land outside of the floor, shake and dump them again.
4. The Simulator leaves in place any beans that are more than halfway inside an infected room, and returns all others back to your hand/cup. For each bean left, find the closest uninfected room to infect and mark it with an 'X'. Afterwards, add these beans to the infected pile.
5. The Recorder counts those the Simulator left behind, tallies this number (Δx) under *Change in Infected*, and then updates the table for the next round by subtracting Δx from the number susceptible and adding Δx to the number infected.
6. Repeat steps 3-5 until there are no more susceptible beans left. When complete, erase all marks from the floorplan (or overhead slide strip).

Model Development

As a group, complete the following. This is most easily done on a single computer when the group is together. Otherwise, work on a shareable spreadsheet (such as Google Sheets).

1. Type all ten simulation tables into a spreadsheet, each table expanded as show below. The *Change in Infected* is a rough estimate for the derivative (rate of change in the infected population), and can be improved by taking the average of the current and previous values to compute the *Infected Derivative*. Note, this means the leftmost value (at $t = 0$) is $\frac{0+\Delta x}{2}$. Once the group model is chosen, its solution is computed in the *Model: Infected* row and square errors are computed below it and summed up.

Simulation #	Parameter values			
Time (rounds)	0	1	2	...
Data: Susceptible	$N - x_0$			
Infected	x_0			
Change in Infected	Δx			
Infected Derivative	$\frac{0+\Delta x}{2}$			
Model: Infected				
Square Errors				

2. Choosing a simulation with the lowest starting value x_0, create a scatter plot with independent variable *Time* and dependent variables *Susceptible, Infected, Infected Derivative.*
3. From this scatter plot, individually come up with a possible differential equation model to match the shape of the slope $\frac{dx}{dt}$ (Infected Derivative) using x (Infected), $N - x$ (Susceptible), and t (Time). The instructor may have some suggestions to choose from or leave it open-ended.
4. Get together as a group and compare models. Graph/sketch the slope functions from each and work together to choose/create the model with the most potential. Make sure there is at least one parameter to provide room to fit it to the data.
5. Summarize why your group's model fits the shape of the data and how the parameters adjust it. Briefly share this with the class and record any feedback. Update your model accordingly.

Analysis

Your analysis begins with fitting your model equation using a trend line and then a hands-on optimization with the main parameter(s) using either a local search or global search method.

Parameter Estimation

1. Determine the model component X to needed to write your model differential equation as $Y = a_1X + a_0$ (linear) or $Y = a_nX^n + ... + a_1X + a_0$ (polynomial). *For example,* $x' = a \sin bx$ *is linear with* $X = \sin bx$ *where b can be estimated from the period of the oscillation in the data.*
2. Choosing an example simulation with the lowest starting value x_0, insert a row labeled *Model component:* X below the *Infected* row and compute the model component, X.
3. Create a scatter plot with independent variable *Model component:* X and dependent variable *Infected Derivative.* Add a trendline and edit it to use the correct form and display the equation and R-squared value on the chart. Also, set the intercept if you have no constant to solve for. This is your initial estimate for the parameter(s). Note, the R-squared value gives the percentage of the spread in the data explained by a linear relationship and is one way of measuring model fit.
4. Solve your model for the general solution.
5. For each simulation, enter the general solution into the spreadsheet row *Model: Infected* using your initial estimate for the parameter(s) from this example simulation. Make a scatter plot with independent variable *Time* and dependent variables *Infected, Model: Infected*
6. Complete either **Local Optimization** or **Global Optimization** section to optimize model parameter estimates for each simulation. Note how the graphs of the model and data are brought closer together with each improved estimate.
7. Summarize the values for the fitted parameter(s) for all ten simulations by the overall mean, and interval [min, max] for each parameter that was estimated this way. If another method was used, such as estimating the period of oscillation make a note of this.

Local Optimization

1. For each simulation, compute (Infected−Model:Infected)^2 in the spreadsheet row *Square Errors* and sum up the row to compute the sum of square errors: $SSE = \sum(Data - Model)^2$.
2. Choose the initial interval radius, Δp, for midpoint $M = p$ and compute the end points $L = p - \Delta p$ and $R = p + \Delta p$. As a rule of thumb, start with $\Delta p = 10^k$ where k the power of ten when p is written in scientific notation.
3. Compute the SSE at end points L and R.
   If $SSE(L) < SSE(R)$, then reset $R = M$ and $M = \frac{L+M}{2}$ to choose the left sub-interval.
   Else, reset $L = M$ and $M = \frac{M+R}{2}$ to choose the right sub-interval.
4. Repeat steps 2-3 until you have obtained an M value with three significant digits unchanged from the previous M value. Set $p = M$.

Global Optimization

1. For each simulation table, copy the pair of rows, *Model: Infected* and *Square Errors*, and paste this pair ten times below it.
2. For a chosen parameter with initial estimate p, choose an initial radius $\Delta p = \frac{1}{2}p$ of your search interval. If there are any other parameters, use their initial estimate value.
3. Divide the interval $[p - \Delta p, p + \Delta p]$ into ten sub-intervals and assign each of the eleven points to a different model pair.
4. For each model pair, compute (Infected−Model:Infected)^2 in the spreadsheet row *Square Errors* and sum up the row to compute the sum of square errors: $SSE = \sum(Data - Model)^2$. Choose the parameter value p^* with minimum SSE.
   If p^* is the same as p, reset $\Delta p = \frac{1}{10}\Delta p$ to zoom-in, and repeat steps 3-4 (unless this makes $\Delta p < 1E - 14$, in which case go to step 6).
5. If at least three significant digits of p^* agree with p, then update $p = p^*$ for the final estimate. Otherwise, repeat steps 3-4 with $\Delta p = |p - p^*|$ and $p = p^*$.
6. Repeat this process for each additional parameter.

In-Class Presentation

1. Explain the main steps in deriving your model differential equation
2. Explain the main steps needed in solving your model differential equation
3. From the ten simulations, show example graphs of your best and worst fitting models to the data.
4. Explain in what ways your model described the data well and in what ways did it not.
5. How could your model be improved? How could the simulation of the spread of a cold be improved?

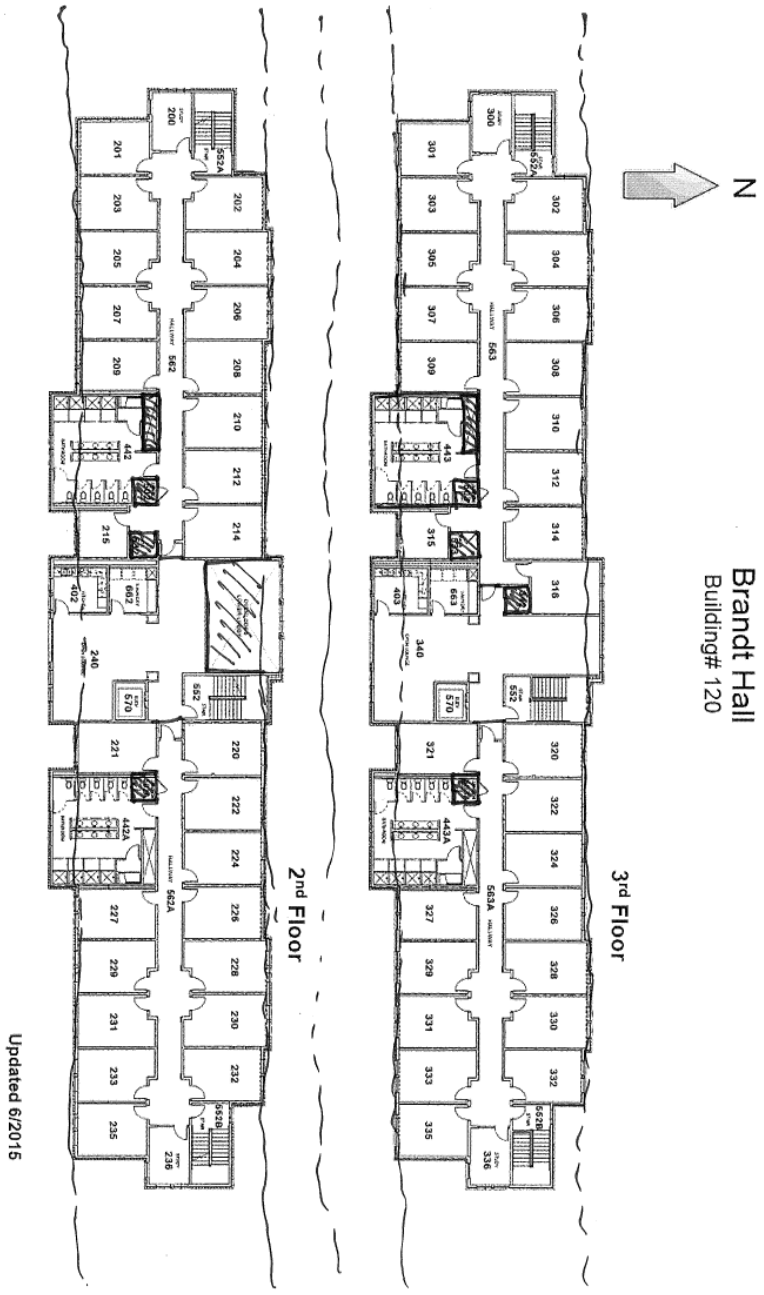

Figure 1. Example floor plan handout